

混合整数計画によるパターン認識のモデル化と学習 アルゴリズム

松永, 浩之

<https://doi.org/10.11501/3134562>

出版情報：九州芸術工科大学, 1997, 博士（工学）, 課程博士
バージョン：
権利関係：

氏名・本籍(国籍) 松永浩之(福岡県)
学位の種類 博士(工学)
学位記番号 甲第15号
学位授与の日付 平成10年3月18日
学位授与の要件 学位規則第4条第1項該当
学位論文題目 混合整数計画によるパターン認識のモデル化と学習アルゴリズム

ム

審査委員会 幹事 教授 浦濱喜一
委員 教授 瀧山龍三
委員 教授 長島健次
委員 教授 福島重廣

論文内容の要旨

生体における知覚は感覚入力に基づいて外界の状態を推定する情報処理過程であり、広い意味でパターン認識の問題として捉えられる。認識システムの入出力関係はノイズや遮蔽等の影響により一般に多対-写像となっているので、その逆写像の解は複数あり得る。従って、認識システムはそれらの解の候補の中から最も確からしい解を選択して出力していると考えられる。そこで最適化問題によって定式化を行なった視知覚のモデルがこれまでに数多く提案されてきた。また、一般に生体のパターン認識はある物体が何であるかというカテゴリー化を行なう離散的な処理とその物体の大きさなどの物理量を推定する連続的な処理という性質が異なる二つの処理過程が密接に関係しあっている。しかしこれまでに提案されたモデルのほとんどは連続的な処理か離散的な処理のどちらか一方しか扱っていない。本論文では、これら両方の処理過程を不可分に扱う視知覚の数理モデルを混合整数計画問題によって定式化する。

本論文は、知覚や学習の混合整数計画問題による数理モデル化について研究した結果をまとめたものであり、8章から構成されている。

まず第1章では、本研究の背景と扱っている問題を示し、あわせて論文の概要について述べる。

次に第2章では知覚や学習を混合整数計画問題によってモデル化する際に必要となる数学的基礎を概括し、技術的な問題を論じる。ここでは、生体の情報処理過程をパターン認識システムとして捉え、混合整数計画問題として定式化をおこない、認識システムの学習について議論し、認識や学習の具体的な定式化を最近傍識別を例にとり説明する。また混合整数計画問題の解法として勾配系を用いて解くアナログ解法を提案する。

第3章以降では、この基礎理論を知覚や学習の問題に適用する。第3章と第4章では知覚を扱い、第5章、第6章、第7章では学習を扱う。続く第8章で第3章から第7章までの結果を統合して、マルチモーダルパターンの認識モデルと学習法を提案する。

第3章ではラインプロセスを拡張した結合プロセスを提案し、画像処理への応用を示す。

ラインプロセスではしきい値が絶対的な値に固定されるので、グレイレベルのスケール変化によって結果が変わるという欠点がある。結合プロセスはこの欠点を緩和し、エッジを保存して平滑化を行なうことができる。またパラメータを変化させることによってマルチスケール表現が得られる。更にインパルスノイズの除去もできるように拡張し、セグメンテーション等も容易に得られることを計算機実験により示す。

第 4 章では結合プロセスが生体の視知覚のモデルとして有効であることを示すために、結合プロセスを用いて幾何学的錯視のモデル化を行なう。パターンの知覚体制化には様々な要因があるが、ここではドットパターンを対象として最も基本的であると考えられる近接要因による群化のモデルを提案する。代表的な幾何学的錯視図形に対して計算機実験を行ない、錯視現象のモデルとしての結合プロセスの有効性と限界を示す。

第 5 章ではクラスタリング関数回帰による最近傍識別器の教師あり学習法を提案する。この方法は教師信号に基づいて定数関数への回帰を学習することによって、最近傍識別器の代表点の最適な配置を決める方法である。この方法を用いると代表点の数を多めに準備しておけば必要な分だけが割り当てられることや木探索識別器の代表点を一括して最適化する学習が行なえることを計算機実験により確認する。

第 6 章では第 5 章で提案した方法を線形射影を用いてデータベクトルの次元の圧縮を行なうように拡張し、その有効性を実験により示す。教師なし学習ではデータの復元誤差が最小となるように、教師あり学習では識別誤差が最小となるように最適化問題として定式化する。簡単なデータを用いた計算機実験により本学習法の特性を明確にするとともに、パラメータ化された顔データの分類に適用する。

第 7 章では教師ありのデータと教師なしのデータが混在する場合の最近傍識別器の学習法を提案する。ここでは、代表点の配置を決定するのに識別誤差と量子化誤差を共に最小化するように最適化問題として定式化する。更に教師ありデータだけで学習するよりも教師なしデータも付け加えて学習する方が識別率の期待値が向上することを、理論的な解析と顔データを用いた計算機実験とによって実証する。

第 8 章ではロバスト推定とベイズ則に基づく複数の入力からなる異種データの統合すなわちマルチモーダル情報による最近傍識別法を提案し画像処理への応用を示す。この統合方法は多数決をファジイ化したものであることを示し、EM アルゴリズムによる教師なし学習法を導出する。更にこの識別方の基本的な性質を簡単なデータで検証し、視覚と聴覚の統合の例としてマガーク効果を定性的に説明する。

最後に第 9 章では、本研究のまとめと今後の課題を述べる。

論文審査の結果の要旨

脳におけるパターン認識の情報処理様式の研究は古くから行われてきており、生理学や心理学の進展に伴って数理モデルの研究も活発化してきている。このような生理・心理学的知見に基づくパターン知覚の数理モデルはコンピュータビジョンなどの工学的な応用への基礎となると期待される。これまでのパターン認識の手法の多くは記号処理に代表され

るデジタルな手法とニューラルネットなどのアナログ主体の方法とに2分されているが、総合的な性能の向上が望まれており、脳ではデジタルアナログの両処理が有機的に結合してロバストなパターン知覚が行われているという生理・心理学的知見から、両処理の統合化に期待が寄せられている。

本論文では、脳でのパターン知覚がデジタルとアナログの統合処理であること、最適化問題の求解プロセスとして記述可能なことに基づき、パターン認識を混合整数計画問題のアナログ解法としてモデル化している。またパターン認識器の学習も同様に混合整数計画問題として定式化してパターン知覚の心理実験結果のシュミレーションや工学的な応用例を示している。

第1章では、脳のパターン認識の様式に関する知見を概括し、混合整数計画問題による数理モデルの論拠を提示している。

第2章では、混合整数計画問題のアナログ解法を示し、パターン認識システムの定式化に加えて、その学習のアルゴリズムを提示している。

第3章では心理学や生理学で存在が示唆されている結合プロセスと呼ばれるニューロン間の相互作用をモデル化し、ノイズの平滑化やマルチスケールの画像復元やセグメンテーションに有効であることをシュミレーションで実証している。

第4章では、前章で提示した結合プロセスモデルが視覚パターン知覚のモデルとしても有効であることを幾何学的錯視のシュミレーションで示している。ドットパターンを見たときに知覚される幾何学的な変位をクラスタリングに基づいてモデル化し、ミュラーリヤー図形を代表例として数多くの錯視図形の心理実験結果を半定量的に再現できることを示している。

第5章では、最も簡単で代表的なパターン認識法である最近傍識別器の新しい学習法として、クラスタリングと関数回帰とに基づく方法を提案し、各カテゴリーへ代表点が自動的に割り当てられ、木探索識別器の一括学習により性能が向上するなどの優れた性能を実証している。

第6章では、前章と同じく最近傍識別器を扱い、線形射影によるデータの次元圧縮を導入した学習法を提案し、顔の識別に適用して有効性を実証している。

第7章では、引き続き最近傍識別器について、教師ありデータと教師なしデータが混在している場合の学習法を提案している。パターン識別器の学習においては専門化が予めラベル付けしたデータが用いられるが、多数のデータへのラベル付けは労力が多く困難である。そこでラベルなしのデータも組み合わせた効率的な学習法が望まれている。ここで提案した方法はそのような学習法の一つであり、理論および顔画像認識の実験で有効性を実証している。

第8章では、画像と音声などマルチモーダルなパターンの認識システムとその学習アルゴリズムを提案している。多数決をファジー化した統合様式を提案し、マガーク効果の心理実験結果を再現し、多重画像のセグメンテーションやノイズ除去に有効であることを実

験で示している。

最後に第9章では、得られた成果を総括し、今後の課題について検討を加えている。

以上のように、本論文はパターン認識の混合整数計画による統合的な数理モデルを提案したものであり、心理実験結果を再現しているとともに工学的な応用例も提示されており、今後コンピュータビジョンなどの研究開発に寄与しうると考えられ、工学博士の学位論文に値するものと認められる。

最終試験の結果の要旨

本論文の全般にわたって著者の説明を受けた後、審査委員が口頭により試問を行なった。

(1)ファジー化の根拠、(2)マルチモードと多次元特徴量との差違及びモードの独立性の仮定の妥当性、(3)混合学習でのパラメータの設定法、(4)ノイズ平滑化と画像復元及び領域分割とセグメンテーションとの関係、(5)認知心理学的モデルと工学的応用との関連性、などについて質問があり、何れに対しても明快な回答が得られた。

次に本論文について情報伝達専攻主催の公開発表が開かれ、本学内研究室、他大学及び地元企業等から多数の出席者があった。著者の発表の後の質疑討論では、認知心理モデルとしての有用性、マルチモーダルパターン認識の工学的応用、画像処理応用への今後の展開などについて活発な質問がなされ、著者の説明によっていずれに対しても十分な理解が得られた。

以上の結果から、著者は最終試験に合格したものと認定した。

目次

第 1 章	序論	5
1.1	研究の目的と背景	5
1.2	論文の概要と構成	9
第 2 章	混合整数計画問題	13
2.1	まえがき	13
2.2	パターンに対する知覚	13
2.3	認識システムの学習	16
2.4	最近傍識別則	16
2.5	最近傍識別器の学習法	17
2.6	混合整数計画問題の解法	18
2.7	むすび	22
第 3 章	結合プロセスによる画像の復元とセグメンテーション	24
3.1	まえがき	24
3.2	ラインプロセスによる画像復元	25
3.3	結合プロセスによる画像の復元とセグメンテーション	26
3.3.1	画像復元	28
3.3.2	インパルスノイズの除去	28
3.3.3	セグメンテーション	31
3.4	マルチスケール表現	31

3.4.1	連続変形によるマルチスケール表現	31
3.4.2	拡散によるマルチスケール復元	36
3.5	スパースなデータへの応用	36
3.5.1	スパースデータからの全画像の復元	36
3.5.2	ドットパターンの復元とセグメンテーション	38
3.6	むすび	38
第 4 章	結合プロセスによるドットパターンの群化モデル	43
4.1	まえがき	43
4.2	結合プロセスモデル	44
4.3	シミュレーション結果	46
4.3.1	ドットパターンの群化	47
4.3.2	錯視	47
4.4	むすび	55
第 5 章	クラスタリング回帰による最近傍識別器の学習法	69
5.1	まえがき	69
5.2	最近傍識別器	70
5.2.1	教師なし学習	70
5.3	クラスタ分割関数回帰	71
5.3.1	クラスタリング回帰	72
5.3.2	階層クラスタリング回帰	73
5.4	最近傍識別器の教師あり学習	74
5.4.1	代表点の割り当て	75
5.4.2	関数の次元	77
5.5	むすび	79
第 6 章	アフィン部分空間への射影による最近傍識別器の学習法	82
6.1	まえがき	82

6.2	アフィン射影最近傍パターン識別器の学習	83
6.2.1	教師なし学習	83
6.2.2	教師あり学習	89
6.3	むすび	97
第7章	最近傍識別器の教師あり / なし混合学習	102
7.1	まえがき	102
7.2	教師あり / なし混合学習	103
7.3	実験例	106
7.4	むすび	108
第8章	マルチモーダルパターン識別器の教師なし学習	115
8.1	まえがき	115
8.2	最尤推定による識別と学習	116
8.3	確率分布の例	117
8.4	識別の統合則について	118
8.5	学習例	120
8.6	多重画像のセグメンテーション	135
8.7	むすび	136
第9章	結論	141
	謝辞	145
	参考文献	146

第 1 章

序論

1.1 研究の目的と背景

生物は日常活動する中で様々なものを知覚し、高度な情報処理を行っている。生物にとって知覚は必要不可欠で重要な機能であり、この機能が無ければ生存していくことは不可能である。知覚は常に無意識に行われていることであるが、計算機で実行することは極めて困難な課題である。産業界における機械化や自動化の流れの中で、より柔軟で人間的な計算機の要求が高まっており、このような知覚における脳や認知の情報処理過程を理解することは時代の要請である。認知機能を理解することによって、個人識別、文字認識、音声認識、医用画像処理、リモートセンシングなど多くの分野への応用が開けると期待される。一方、計算機の処理速度、記憶容量は日進月歩であり単純な計算であればかなり大規模な問題が解けるようになってきた。その計算機の能力を利用して、物理学の分野では多粒子系の問題など複雑な過程の数値計算による実験が盛んに行われ成果を上げている。理論によって構築されたモデルを計算機で実行することによって、そのモデルの持つ定性的な理解を促したり、実験の結果をモデルや理論にフィードバックして修正を行うことが行われている。このようなことから、本論文では基礎的な知覚や学習の数理モデルを構築し、計算機実験を行うことによって、それら情報処理過程を定性的に理解し、画像処理やパターン認識等の分野への応用を示す。

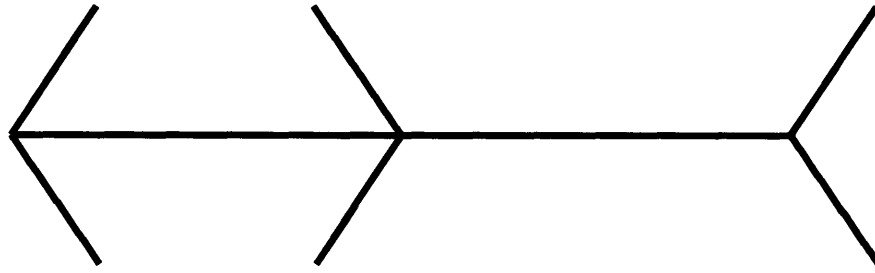


図 1.1: Muller-Lyer 図形

知覚は心理学・生理学・工学等の幅広い分野で研究されている。これまで心理学は主に人間の認知過程の実験による測定や現象の記述といった巨視的な過程を問題としてきた。錯視研究の場合、被験者に対して様々な比較刺激が提示されパラメータと錯視量の関係等が定量的に測定される。例えば図 1.1に示す Muller-Lyer 図形の場合、矢線の長さや角度を変化させた図形が示され、それによってどのくらい違う長さが知覚されるかが測定されその関数関係が記述される。その際、どのようにパラメータをとり何を測定するかは作業仮説に依存しどのような仮説を立てるかは大変重要な問題である。本論文では錯視のモデルを示すが、ここでは錯視が起こる要因として体制化を仮定した。体制化の要因には近接性、類同性、連続性、対象性、閉合の法則 [1] があるが近接性だけを仮定してモデルを作成している。その理由は、できるだけ単純なモデルが良い事、その単純なモデルでどこまで錯視が説明できるか調べる事を目的としたからである。また心理学では、物理的に存在しない線を知覚する主観的輪郭の知覚 [2] についても研究されている。そこでは、神経回路網を意識したモデルも提案されている [3, 4] が繁雑でありその動作に対する見通しが立てにくい等の難点がある。本論文の後半で議論するカテゴリー化についても心理学で研究されており、代表的なものとしては2種類のモデルが提案されている。一つは事例モデル [5] と呼ばれ個々の事例がどのカテゴリーに属するかといった情報とともに記憶される。新奇な刺激に対してはそのつど全ての記憶との類似性からカテゴリーが決定される。このモデルはデータベースに例えることができよう。二つめは原型モデル [6, 7] と呼ばれ、学習を通じてカテゴリーの原型が抽出されその代表のみが記憶される。新奇な刺激に対してはその原型との類似性からカテゴリーが決定される。どちらが人間の認知過程をより良く

反映しているかについては様々な議論があるが、本論文では原型モデルを採用している。理由としては、第一に判断に要する時間の速さや記憶容量の少なさ等の最適性、第二に神経回路モデルとしての表現の親和性があげられる。

次に生理学は主に、神経単体やモジュール、また最近では神経伝達物質等の分子レベルの処理などの微視的な過程を問題としている。生理学における脳の研究では、脳内のどの領域がどのような処理に関わっているかが調べられる。視覚に関わっている視覚野と呼ばれる領域は、その構造や刺激選択性から細かく分類されている。例えばV2野の細胞は、方位、運動方向、色、コントラスト等に対する刺激選択性があり、主観的輪郭にも反応する [8]。本論文では、ニューロン単体の振る舞いを積み上げて巨視的なモデルを組み立てることはしない。巨視的/微視的なアプローチの中間レベルでの数理モデル化を行なう。すなわち、知覚や学習といった巨視的な過程のモデル化を行うが、ニューラルネットといった微視的な背景も考慮する。

工学では、コンピュータビジョン、パターン認識等の研究分野で計算機による知覚の研究が行われている。コンピュータビジョンの分野では、画像入力からの三次元形状の復元や位置の特定の問題を扱う。またパターン認識では、入力データの 카테고리への対応付けが主たる関心事である。そこでは「機械に外界の認識を行なわせるためにはどうすればよいか」が第一の課題として設定されており、生体がどのような処理を行なっているかは問題とされず、その課題を実行するために計算機にとって扱いやすい高速なアルゴリズムが要求される。本論文で扱う問題は広い意味でパターン認識の問題であるが生体の情報処理に基づいた視点からアプローチする。

一般に生体におけるパターン認識は、外界からの感覚器に対する刺激に基づいて真の属性や状態を推定する情報処理過程である。この時認識システムへ与えられる外界からの入力、様々な変換を受けたりノイズが加わったりしている。ある物体を見た場合、まずそこで得られる入力は物体に反射した光であり、匂いや硬さ等の情報は欠落している。更に網膜上に映った像は三次元物体が二次元に射影されたものであり遮蔽等も起こっている。元の実在する物体はかなりの変換や擾乱が加わって視覚野へと伝えられる。視知覚の目的は、この実在する物体の部分的な情報である二次元像からその物体の様々な属性、物体までの距離、物体の動く方向や速さ等を推定するこ

とである。このように部分的な情報から推定するのであるから、一般に解は一意ではなく多数存在する。しかし、人間の意識に上る結果は一時期に一つである。例えば有名な「ルビンの壺」の図を見た時、それが向かい合っている顔に見えるか壺に見えるかのどちらかであり両方同時に意識されることはない。従って、人間の知覚システムは多数ある解の候補の中から最も確からしいものを一つ選択して出力していると考えられる。すなわち、選択する基準が評価関数で与えられるとすると、知覚システムは最適化問題として定式化できると言える。最適化問題は一般に数理計画問題と呼ばれ、解の取り得る値の範囲によって、整数計画問題 [9] と非線形計画問題 [10] とに分類される。最適解の取り得る値の範囲が整数であるものと実数であるものが混在するものを混合整数計画問題という。

次にこれらの最適化問題の中からどれが知覚のモデル化に適しているかを考える。人間が知覚システムの出力として得られる結果は離散量と連続量の二つの場合がある。例えば“a”という文字と“d”という文字を見た時、それが手書き文字で“a”と“d”の中間的な文字であっても、必ずどちらかに決められるので視知覚システムから出力される結果は離散量である。但しここでは文脈や単語といった高次の処理については扱わず、低次の知覚レベルの処理を問題としていることに注意されたい。またボールが自分に向かって飛んで来る時は、物体までの距離、物体の動く方向や速さ等によって自分にぶつかるまでのタイミングを判断しており、これは連続量である。このように知覚システムには連続的な処理と離散的な処理の二つの処理が混在しているが、これらは密接に関連し合って遂行していると推察される。ここでは例として画像のセグメンテーションすなわち遮蔽やテクスチャーのある面を画像の中から切り出す問題を考えてみよう。テクスチャーのある滑らかな面を一つの面として認識するには、そのテクスチャーをひとつのカテゴリとして判断する離散的な処理とテクスチャーのある面の勾配を推定する連続的な処理の両方が同時に行わなければならない。また遮蔽のある二枚の面があるとする、ここでセグメンテーションを行うには一つの面の連続性ともう一方の面との不連続性が同時に問われる。一般に複雑な背景から一つの図を浮かび上がらせると言うことは、その図のひとまとまりとしての連続性と他のパターンとの不連続性を認識することであり、そこでは連続的な処理と離散的な処理の両方

が不可分に行われている。画像復元のモデルとしては標準正則化理論に基づく RBF ニューラルネット [11] やマルコフランダム場モデル [12], パターン認識としては RBF ニューラルネット [11] やホップフィールドネット [13] など最適化問題による視覚のモデル化はこれまでも多く提案されている, しかしこれらのモデルは連続的処理か離散的処理のどちらか一方しか行っていないので知覚のモデルとしては不十分であると考えられる。そこで本論文では連続的処理と離散的処理を混在させた混合整数計画問題による視知覚のモデルを作成しその有効性を検討する。非線形計画問題などと違い, 整数計画問題は組合せ最適化問題を定式化したものであり NP 困難な問題が多い。従って混合整数計画問題も一般に解を求めるのが難しい。このような問題の解法としては分枝限定法 [14], 動的計画法 [15], 遺伝的アルゴリズム [16] などといったいずれもデジタルな解法が提案されている。ここで問題としているのは生体のニューラルネットを背景とした知覚のモデルであり, それはアナログシステムだと言われている。従って本論文では混合整数計画問題のアナログ解法を提案する。

1.2 論文の概要と構成

まず第 1 章では, 本研究の背景と扱っている問題を示し, あわせて論文の概要について述べる。

次に第 2 章では視覚パターンの知覚や学習を混合整数計画問題によってモデル化する際に必要となる数学的基礎を概括し技術的な問題を論じる。本論文では, 生体の情報処理過程をパターン認識システムとして捉え, 認識システムの出力が最適化問題の最適解として与えられると仮定する。認識システムの出力はカテゴリーのように整数の時もあれば物理量のように実数の時もあるので, 認識システムは混合整数計画問題として定式化できることを述べる。また, 認識システムの学習についても混合整数計画問題として扱えることを示す。これら認識や学習の定式化を最近傍識別を例にとり具体的に説明を行う。更に認識システムとしての混合整数計画問題の解法として勾配系を用いて解くアナログ解法を提案する。このアナログ解法は生体のニューラルネットのモデルとして捉えることができる。

第3章以降では、この基礎理論を知覚や学習の問題に適用する。第3章と第4章では知覚を扱い、第5章、第6章、第7章では学習を扱う。続く第8章で第3章から第7章までの結果を統合して、マルチモーダルパターンの認識モデルと学習法を提案する。

第3章では心理学や生理学の分野で linking process[17] とか associative field[18] などと呼ばれているニューロン結合を数理モデル化したものとして結合プロセスを提案し、画像処理への応用を示す。ラインプロセス [19] ではしきい値が絶対的な値に固定されるのでグレイレベルのスケール変化によって結果が変わるという欠点がある。結合プロセスはラインプロセスを拡張したものとしても捉えられ、この欠点を緩和しエッジを保存して平滑化を行なうことができることを計算機実験により示す。また結合プロセスをインパルスノイズの除去もできるように拡張し、この方法がガウスノイズの除去においてはメディアンフィルタより優れていることやスパースなデータからの復元やセグメンテーションを容易に行なえることを計算機実験により確認する。更に、結合プロセスの式に与えるパラメータを変化させたり、非線形拡散方程式を用いることによりマルチスケール表現が得られることを示す。

第4章では結合プロセスが生体の視覚パターン知覚のモデルとしても有効であることを示すために、結合プロセスを用いた視覚パターン知覚のモデルを提案し幾何学的錯視の計算機実験の結果を示す。パターンの知覚体制化の要因には近接性、類同性、連続性、対象性、閉合の法則があるが、本論文ではドットパターンを対象として最も基本的であると考えられる近接要因による群化のモデルを提案する。このモデルを作成するにあたっては、結合プロセスによって心理的なドットの変位が起こり錯視のようにパターンが変形すると仮定し、心理実験の結果である非線形性とスケール不変性の性質を前提とした。Muller-Lyer 図形を対象として計算機実験を行ない変位量の変化や不動点の位置が心理実験の結果を再現していることや矢線の角度や長さの変化による錯視量の変化について心理実験の結果と定性的に一致することを示す。更に同心円錯視、分割距離錯視、曲率の錯視や代表的な幾何学的錯視図形に対して計算機実験を行ない錯視現象のモデルとしての結合プロセスの有効性を確認する。

第5章ではクラスタリング関数回帰による最近傍識別器の教師あり学習法を提案す

る。従来提案された様々な学習法には、各カテゴリーの代表点の数を予め設定しなければならず、また木探索識別器には適用できないなどの欠点がある。そこで本論文ではこれらの難点を解消する方法として関数回帰に基づく学習法を提案する。この方法は識別率が最小となるように最適化問題として定式化して教師信号に基づいて定数関数への回帰を学習することによって、最近傍識別器の代表点の最適な配置を決める方法である。この学習法を用いると代表点の数を十分多く準備しておけば、各カテゴリーに必要な最小限の数の代表点が割り当てられ、不要な代表点はどのカテゴリーにも割り当てられずに余ることが計算機実験によって示す。またこの学習法を木探索識別器の代表点を一括して最適化する学習が行なえるように拡張する。二段木探索クラスタリングにおいて回帰する関数の次元が学習の効率に与える影響を調べ、次元数が低い場合には一括学習の方が逐次学習より良い結果が得られることを計算機実験により示す。

第6章では第5章で提案した方法の特徴抽出すなわちデータベクトルの次元の圧縮を行なうように拡張しその有効性を実験により示す。次元の圧縮には最も基本的であるアフィン部分空間への線形射影を用いる。この場合カテゴリーの代表が点ではなく超平面で表され、入力データを平面へ射影した残差が最も小さいカテゴリーに分類する。本論文では、教師なし学習では射影したデータの復元誤差が最小となるように最適化問題として定式化し、教師あり学習では第5章で提案したクラスタリング関数回帰による方法を用いる。部分空間の取り方として、射影行列がクラス毎に異なり原点は同じ、射影行列が同じで原点がクラス毎に異なる、射影行列も原点もクラス毎に異なる、の3通りを考えそれぞれに対して簡単なデータを用いて実験することによりその特性を明確にする。更にこの方法によりパラメータ化された顔データの分類を行ない高い識別率が得られることを計算機実験により示す。

第7章では教師ありのデータと教師なしのデータが混在する場合の最近傍識別器の学習法を提案する。学習後の識別器は、教師あり学習では専ら識別に使われ、教師なし学習では識別に使われることもあるし、ベクトル量子化に使われることもある。これらの学習はどちらも代表点のある種の最適な配置を求めるものであるが、最適化の目的関数が異なるので得られる配置は一般的に異なり、一方の学習結果は他方の目的にとっては最適とはならない。ここでは、代表点の配置を決定するのに識別誤差と量

量子化誤差を共に最小化するように最適化問題として定式化する。この方法では荷重係数を変化させると量子化誤差と識別誤差が単調に変化することを理論的に示す。更に教師ありデータだけで学習するより、教師なしデータも付け加えて学習する方が識別率の期待値が向上することを、理論的な解析で予想する。簡単なデータと顔画像データに対し高次自己相関関数により 25 次元に落したデータを用いた計算機実験によってこの予想を実証する。

第 8 章ではロバスト推定とベイズ則に基づく複数の入力からなる異種データの統合すなわちマルチモーダル情報による最近傍識別法を提案する。この方法は多数決をファジイ化したものであることを示し、EM アルゴリズムによる教師なし学習法を導出する。この識別法の基本的な性質を簡単なデータで検証し、視覚と聴覚の統合の例としてマガーク効果を定性的に説明する。更に多重画像のセグメンテーションの計算機実験を行い応用例を示す。

最後に第 9 章では、本研究のまとめと今後の課題を述べる。

第 2 章

混合整数計画問題

2.1 まえがき

知覚は外界からの入力に対する情報処理過程であり，生物におけるその処理速度から推察すると，高速に効率的な処理が行われていると考えられる．本章では，知覚や学習等の視覚情報処理の問題が混合整数計画問題として定式化できることを述べる．効用の度合を表す関数が最小または最大となるような変数の値を与えられた条件を満たすように求める問題を数理計画問題という．混合整数計画問題は数理計画問題の一種であり，求める解の値が整数であるものと実数であるものが混ざっている最適化問題である．これらの方法は意志決定や問題解決手法の一つの方法であるが，見通しの立て易さや厳密な解が得られることから広く利用されている．更に本章では混合整数計画問題の勾配法による解法であるアナログ解法を示す．本論文は生体の情報処理過程の数理モデル化を目指すものであり，アナログ解法はその意味で他の手続き的な解法より神経回路に基礎をおく生体の情報処理モデルとして適している．

2.2 パターンに対する知覚

知覚は外界からの刺激として入力情報を受取り，そのパターン刺激の属性の判定結果を出力する情報処理過程である．刺激として感覚器に与えられた情報は，まず神経

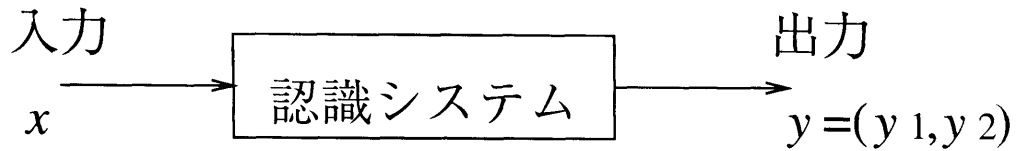


図 2.1: パターン認識システム

パルスに変換されて伝達される。また、それらの情報は細部にわたって全てが伝達される訳ではなく効率化のために特徴抽出等の圧縮が行われる。それとは逆に、外界からの情報に欠落があったり雑音が交じったりしていても補うことができる。これらの処理は利点ばかりではなく、知覚にとっては欠点となる現象である錯視の原因となることもある。このように情報処理過程は入力に対して、それをそのまま高次の処理過程に伝えるのではなく、変換、圧縮、補間等の様々な積極的な処理が行われていると考えられる。本論文ではこれら知覚の情報処理過程の中から重要と思われるいくつかの処理について数理モデル化することを試みる。数学的な取り扱いのために、入力はパターンとして与えられるとする。すなわち、ここでは知覚の情報処理過程をパターン認識の問題としてとらえる。パターン認識とは、図 2.1 に示すような信号等が入力されたときその信号のカテゴリーや物理量として表される属性を出力する処理のことである。この図 2.1 で x は入力パターンであり、実数ベクトルで表される。すなわち入力パターン x は n 次元実数空間 R^n の中の 1 点として表される。一方出力 $y = (y_1, y_2)$ は一般に k 次元整数空間と m 次元実数空間の直積空間 $Z^k \times R^m$ の中の 1 点である。整数 y_1 は通常 1 次元で十分である。この時 y_1 はカテゴリー番号であり入力パターン x がどのカテゴリーに所属するかの判定結果である。また出力 y_2 は知覚した結果の物理量を表す。例えば、文字認識の場合は出力は整数 y_1 だけであり何という文字であるかだけが番号で出力される。一方運動知覚の場合は運動している物体が何であるかを示す整数 y_1 と物体の速度を示す実数 y_2 が出力される。一般に知覚は、ある一つの物体に注意を向けてもその物体の名前、色、形等のカテゴリーや大きさ、速度等の物理

量のような多くの処理結果を出力する。

さて次にこれらの処理過程を数理モデル化することを考える。生き物にとって知覚による判断は生存を左右する重大な問題である。知覚は進化のプロセスを経て複雑化し発達してきているのであるから、より生存に有利な方向へと変化してきていると考えられる。例えば頻繁に間違える知覚情報処理を行う生き物より、より正確な判断が行える処理を行う生き物の方がより多く生存する確率が高い。あるいは同じ処理を行うにも処理速度が速く、エネルギー消費が少ない生き物の方がより多く生存する確率が高い。このような理由から、本論文では、このような入出力関係はなんらかの最適化問題の結果として与えられると仮定する。図 2.1 のシステムの入出力関係は数式ではパラメータ p を用いて、 $y = f(x, p)$ と書ける。本論文ではこれを、 x がある値のとき出力 y が

$$\begin{aligned} \min_y \quad & g(y, x, p) \\ \text{subj.to} \quad & y \in S \end{aligned} \tag{2.1}$$

という最適化問題の最適解として与えられ、入力 x の様々な値に対するこの最適解 y の変化が $y = f(x, p)$ という入出力式になると考える。式 (2.1) の最適化問題は一般に数理計画問題と呼ばれ、変数 y が取り得る値によって以下のように分類される。 y が整数値をとるとき整数計画問題 [9] と呼ばれ、実数値をとるとき非線形計画問題 [10] と呼ばれる。また上記のような整数と実数が混在するような一般の場合を混合整数計画問題と呼ぶ。図 2.1 の認識システムにおいて出力が整数変数 y_1 だけである単なる識別の場合は、認識システムは整数計画問題で表される。一方出力が整数 y_1 と実数 y_2 の両方のときは単なる識別より広いパターン認識が行われており、混合整数計画問題となる。なおパターン識別の場合でも処理全体は混合整数計画問題として表されるが、実数出力 y_2 はシステムから出力されず、整数出力 y_2 だけが出力される場合も多い。従ってパターン認識は一般に混合整数計画問題として定式化できると考えられる。

2.3 認識システムの学習

図 2.1 の認識システムで実際に認識が行われるときにはシステムのパラメータ p は固定のままである。このとき各入力 x に対して固定された関数 $f(x, p)$ による結果 y が出力される。そこで、それぞれ異なるパラメータ p の値を持ついくつかの認識システムがあるとすると、この場合、入力 x がすべて同じでもそれぞれの出力は異なり、この中でなんらかの評価基準に基づいて最も優れた y を出力するシステムが望ましいということになる。この問題はシステムの最適化問題であり、入力データの集合が与えられたときのパラメータ p の最適値を求める問題に帰着する。この最適化問題は一般に

$$\begin{aligned} \min_p \quad & \sum_i h(p, x_i, y_i) \\ \text{subj.to} \quad & p \in W \end{aligned} \quad (2.2)$$

と書くことができる。ここで関数 h は出力 y の良さを評価する関数であり、 x_i, y_i は学習用に与えられるデータである。またパラメータ p の取り得る値は実数とする。式 (2.1) の認識問題は一般に混合整数計画問題になると前述したが、ここでは式 (2.2) の学習問題はどのような計画問題になるのかを考える。学習は大きく分けて教師ありの場合と教師なしの場合の二つに大別される。教師あり学習では学習用データとして入力データと出力データの組 x_i, y_i が与えられる。このときには、式 (2.2) は非線形問題となる。また教師なし学習では学習用データとして x_i だけしか与えられない。従ってこの場合は x_i に対する出力 y_i も求めながら p の最適解を求める必要があり、式 (2.2) は混合整数計画問題となる。なお教師あり学習の場合でも中間的な整数変数を求めながら p の最適化を行う場合も多く、このときには教師ありでも混合整数計画問題となる。

2.4 最近傍識別則

これまでに述べたことを最も簡単なパターン識別法である最近傍識別を例に具体的に述べる。最近傍識別とは、入力パターン x とカテゴリーを代表する点との距離を求め、最も近いカテゴリーに分類する規則である。代表点は一般には各カテゴリーに対

して複数あるがここでは簡単のため1つとする。カテゴリが k 個あり、識別器の出力 y を $y = [y_1, \dots, y_k]$ とする。各 y_i は 0,1 変数である。この場合この識別法は次のように表される。

$$\begin{aligned} \min_y \quad & \sum_{i=1}^k y_i \|x - r_i\|^2 \\ \text{subj.to} \quad & \sum_{i=1}^k y_i = 1 \\ & y_i \in \{0, 1\} \end{aligned} \tag{2.3}$$

ここで r_i は第 i カテゴリの代表点ベクトルである。式 (2.3) の最適解では入力パターン x に最も近い代表点 r_i の y_i が 1 になりその他の y_i が全て 0 になるのは明らかである。従って y_i が 1 になったカテゴリ番号 i が識別器の出力となる。識別器の出力 i は整数であるから識別は一般に整数計画問題であるが、このように 0,1 整数計画問題によっても定式化できる。

2.5 最近傍識別器の学習法

まず最初に教師なし学習について考える。この場合には学習用データとしてパターン識別器への入力パターン x しか与えられない。そこで与えられないデータ y については何らかの推定を行う必要がある。ここでは代表点 r による入力パターン x の識別結果と矛盾しないように y を決定することにする。この場合この学習は

$$\begin{aligned} \min_{y,r} \quad & \sum_i \sum_j y_{ij} \|x_i - r_j\|^2 \\ \text{subj.to} \quad & \sum_j y_{ij} = 1 \\ & y_{ij} \in \{0, 1\}, r_j \in R^p \end{aligned} \tag{2.4}$$

という混合整数計画問題となる。これは各入出力パターンは自分に最も近い代表点に所属させられるのでそのときの誤差が最も小さくなるように y と r を決めるということを表している。これはまたクラスタリング問題すなわち学習用データ x_1, \dots, x_q を k 個のグループにグループ分けする問題と等価である。

次に教師あり学習を考える。教師あり学習では q 個の学習用の入力パターン x_1, \dots, x_q が与えられ、これらがどのカテゴリに所属するかという情報も与えられる。すなわ

ち上記の 0,1 変数の正解 $y^* = [y_{ij}^*]$ が与えられる。ここで $y_{ij}^*(i = 1, \dots, q; j = 1, \dots, k)$ は第 i データ x_i が第 j カテゴリーに所属するメンバーシップである。さて第 j カテゴリーの代表点が r_j であるときのデータ x_i の第 j カテゴリーのメンバーシップ y_{ij} は式 (2.3) で与えられ、この y_{ij} は r_j によって変わるから $r = [r_1, \dots, r_k]$ の関数 $y_{ij}(r)$ となる。この $y_{ij}(r)$ の正解 y_{ij}^* が与えられているのであるから最適な r は

$$\min \sum_i \sum_j (y_{ij} - y_{ij}^*)^2 \quad (2.5)$$

の解で与えられる。この解を求めるのが教師あり学習である。なお y は確率ベクトルであるから、より有効な評価関数として式 (2.5) の 2 乗誤差以外にも Kullback-Leibler[20] ダイバージェンスが使える。またこの教師あり学習は次のようにも定式化できる。代表点 r_j の配置とそれらがとる関数値 g_j が与えられると、 r_j のボロノイ領域で値 g_i をとるような区分的定数関数 $g(x)$ によって最近傍識別が行える。

$$\min_{r, g} \sum_i \sum_j y_{ij} (f_i - g_j)^2 \quad (2.6)$$

このときメンバーシップ y_{ij} は

$$y_{ij} = \frac{e^{-\|x_i - r_j\|^2/T}}{\sum_k e^{-\|x_i - r_k\|^2/T}} \quad (2.7)$$

の $T \downarrow 0$ の極限值として与えられる。ここで、 $f_i (i = 1, \dots, q)$ は教師データである。 f_i には、例えばデータ x_i のカテゴリー番号 j を与えればよい。

2.6 混合整数計画問題の解法

ここではこれまでに述べたような知覚や学習のモデルとしての混合整数計画問題の解の求め方について考える。なお整数計画問題はほとんど全て 0,1 整数計画問題によって定式化できるので、ここでは 0,1 整数計画問題だけを考える。数理計画問題の中で解を求めるのが最も簡単なのは制約なしの非線形計画問題であり、この場合変数は実数空間全体の値を取り得るので勾配法などにより局所最適解を求めることができる。また制約条件付きの非線形計画問題もラグランジュ関数の停留点が解となるから同様

なアルゴリズムによって解を求めることができる。整数計画問題は非線形計画問題よりも解を求めるのが難しく、分岐限定法 [14] や動的計画法 [15], さらにシミュレテッドアニーリング法 [21] や遺伝的アルゴリズム [16] など種々の解法が考案されている。本論文は生体のニューラルネットのモデル化を目的としており、生体のニューラルネットはアナログ系であると言われている。すなわちカテゴリーを表す 0,1 整数変数はニューラルネットの中では 0 から 1 までの値をとる実数変数として表されていると考えられている。これは確率あるいはファジィに関連する心理実験などでも確認されている。従って厳密な 0,1 変数を緩和した解法が処理過程で使われている可能性が高い。しかしシステムの最終的な出力はあくまで 0,1 である。例えばこれは文字認識ではあいまいな文字でも最終的にはどれかの文字に決定づけられるからである。このように混合整数計画問題を勾配法で解く方法をアナログ解法と呼ぶことにする。これについてまず変数が 1 個だけの単純な整数計画問題について基本的な手続きを説明する。変数が多い場合には各々の変数について同じ処理を行えばよい。また混合整数計画問題の場合には元々実数である変数には勾配法が直接使えるから、残る整数変数について以下の処理をすればよい。

次式のような 0,1 整数計画問題を解くことを考える。

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subj.to} \quad & x \in \{0,1\} \end{aligned} \tag{2.8}$$

ここで関数 f は上に非凹すなわち上に凸か線形であるとする。本論文で扱う問題はすべてこの条件を満たしている。そうすると式 (2.8) の解と非線形計画問題

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subj.to} \quad & 0 \leq x \leq 1 \end{aligned} \tag{2.9}$$

の解は一致する。なぜなら f が非凹であるから式 (2.9) の最適解は必ず 0 か 1 になるからである。そこで式 (2.9) を勾配法で解けば (2.8) の解を求めることができる。しかしここではこのように勾配法を用いて直接的に解を求める方法ではなく、

$$\begin{aligned} \min \quad & f(x) + T[x \ln x + (1-x) \ln(1-x)] \\ \text{subj.to} \quad & 0 \leq x \leq 1 \end{aligned} \tag{2.10}$$

の $T \downarrow 0$ の極限として式 (2.9) の解を求める方法をとることにする。この極限の求めるには、例えば T を最初大きな値にしておいて $0 < \alpha < 1$ の範囲の適当な定数 α を用いて $T_{step+1} = \alpha T_{step}$ として各 T_{step} で式 (2.10) の解を求めながら T を十分小さな値まで減少させる。この手続きを決定論的アニーリング法 [22] と呼ぶ。このような周りをする理由は学習にある。まず教師なしの学習では上記のクラスタリングの例のように一般に最適化問題は多くの局所最適値を有する求解困難な問題となる。実際クラスタリングは NP 困難な問題である。従って直接勾配法で解を求めてもあまり良い解は得られない。アニーリングを使えば多くの場合大域最適解に近い準最適解が得られることが経験的に知られている [22]。そこで本論文では上に述べたように式 (2.10) において T を最初大きな値にしておき次第に小さくしながら式 (2.10) の解を追跡して $T = 0$ で式 (2.9) の解を得るという方法を用いる。もう 1 つの理由は教師あり学習においては出力が 0, 1 整数では勾配法が使えないためである。従って式 (2.10) のようにファジィ化して整数変数を 0 から 1 までの連続な値をとるように変更して勾配法が使えるようにして同様なアニーリング法によって $T = 0$ で整数解を得る。式 (2.10) の中の $x \ln x + (1-x) \ln(1-x)$ はエントロピーであり、下に凸な関数であるから、式 (2.8) の解は 0 と 1 の間 $0 < x < 1$ になり、 $T \downarrow 0$ の極限で x は 0 か 1 のどちらかになる。統計物理の用語にならって式 (2.10) の T は温度と呼ばれ、 $f(x)$ はエネルギーと呼ばれる。なお本論文に出てくる整数計画問題は式 (2.3) や (2.4) のように、式 (2.8) に制約条件が付け加わった

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subj.to} \quad & \sum_{i=1}^h x_i = 1, x_i \in \{0, 1\} \end{aligned} \quad (2.11)$$

の形の問題である。この式 (2.11) は

$$\begin{aligned} \min \quad & f(x) + T \sum_{i=1}^n x_i \ln x_i \\ \text{subj.to} \quad & \sum_{i=1}^n x_i = 1, 0 \leq x_i \leq 1 \end{aligned} \quad (2.12)$$

というようにエントロピー $\sum x_i \ln x_i$ を加えてファジィ化する。決定論的アニーリングを使う勾配法によるアルゴリズムは式 (2.12) の場合以下のようにになる。まず式 (2.12)

の解は次のようにして求められる．式(2.12)の制約条件 $0 \leq x \leq 1$ を取り除いたラグランジュ関数は

$$L = f(x) + T \sum_{i=1}^n x_i \ln x_i + \lambda (\sum_{i=1}^n x_i - 1) \quad (2.13)$$

であり，このラグランジュ関数の停留点での x は勾配系

$$\frac{dx}{dt} = -\frac{\partial L}{\partial x} \quad (2.14)$$

の収束値として求められる．制約条件 $0 \leq x \leq 1$ を除いた理由は式(2.14)において x の初期値を $0 \leq x(0) \leq 1$ にとっておけば，それ以後 $0 \leq x(t) \leq 1$ は満たされるからである． T の各値でこの方法で x を求めながら， T を大きな値から徐々に小さくして十分小さな T で(2.11)の解を求めるのがこの場合の決定論的アニーリングである．この時の x が式(2.11)の解になる理由は，式(2.10)の時と同じである．次にこの決定論的アニーリングで定数 α を使わずに自動的にスケジューリングする方法を考える．そのために T の初期値に適当な大きな値を入れ

$$\frac{dT}{dt} = \frac{\partial L}{\partial T} \quad (2.15)$$

として式(2.14)と同時に反復収束させる．こうすればスケジューリングを気にすることなくアニーリングが行える．なぜならば式(2.13)から $\partial L / \partial T = \sum x_i \ln x_i$ となり $0 \leq x_i \leq 1$ であるから $\partial L / \partial T \leq 0$ となり，従って(2.15)より $dT/dt \leq 0$ となり T は単調に減少し， $\sum x_i \ln x_i = 0$ となるのは全ての x_i が0か1のときだけであるので，その時 $dT/dt = 0$ となり収束する．この時 $\sum x_i \ln x_i = 0$ であるので式(2.12)は式(2.11)となり， x は式(2.11)の解である．

次に式(2.6)のように教師ありの場合の決定論的アニーリングを使う勾配法によるアルゴリズムは以下のようになる．この場合目的関数は

$$E = \sum_i \sum_j y_{ij} (f_i - g_j)^2 \quad (2.16)$$

であり， y_{ij} には式(2.7)を代入する．この関数の停留点での r と g は勾配系

$$\frac{dr}{dt} = -\frac{\partial E}{\partial r} \quad (2.17)$$

$$\frac{dg}{dt} = -\frac{\partial E}{\partial g} \quad (2.18)$$

の収束値として求められる。上記のように T の各値で r と g を求めながら T を大きな値から徐々に小さくして十分小さな T で式 (2.6) の解を求めればよい。式 (2.16) は y_{ij} に関して 1 次であるから y_{ij} の最適解は許容領域の端点である。式 (2.7) より $0 \leq y_{ij} \leq 1$ なので最適解では y_{ij} は 0 か 1 となる。 y_{ij} が 0 か 1 になるのは $T = 0$ の時であるから、この場合にアニーリングを自動的にスケジューリングするには、 T の初期値に適当な大きな値を入れ

$$\frac{dT}{dt} = -\frac{\partial E}{\partial T} \quad (2.19)$$

として式 (2.17),(2.18) と同時に反復収束させればよい。

式 (2.15) と式 (2.19) の右辺の符号が逆であることに注意されたい。これは、式 (2.13) では T は実はラグランジュ乗数であるのに対し、式 (2.16) では T は最適化変数であることによる。式 (2.13) の T がラグランジュ乗数であることは以下のようにして分かる。式 (2.11) は

$$\begin{aligned} \min \quad & f(x) \\ \text{subj.to} \quad & \sum_{i=1}^n x_i = 1, \sum x_i \ln x_i = 0, 0 \leq x_i \leq 1 \end{aligned} \quad (2.20)$$

と等価である。なぜなら $x_i \in \{0, 1\}$ と $\sum x_i \ln x_i = 0$ は等価であるからである。前述したように制約条件 $0 \leq x_i \leq 1$ を除いた式 (2.20) のラグランジュ関数は

$$L = f(x) + T \sum_{i=1}^n x_i \ln x_i + \lambda (\sum_{i=1}^n x_i - 1) \quad (2.21)$$

となる。式 (2.21) は式 (2.13) と同じである。すなわち T は制約条件 $\sum x_i \ln x_i = 0$ のラグランジュ乗数である。

2.7 むすび

この章では、認識や学習が一般に混合整数計画問題として定式化されることを示し最近傍識別を例に具体的に説明した。また混合整数計画問題の決定論的アニーリング

による勾配法であるアナログ解法とアニーリングのスケジューリングを自動化する方法を提案した。以後、本論文ではこれらの方法を用いて知覚や学習の数理モデルを提示し計算機実験等の結果を示す。第3章と4章は認識システムのモデル化に関するものであり、その後の第5章と6章と7章は学習のモデル化に関するアルゴリズムを議論し、第8章ではこれらの結果を統合してマルチモーダル情報の認識システムと学習のモデル化について述べる。

第 3 章

結合プロセスによる画像の復元とセグメンテーション

3.1 まえがき

正則化による画像復元法において、エッジを保存するようにラインプロセスと呼ばれる変数が導入されている [19]。これについては似たような働きをするニューロンの存在が示唆されたり [23]、アナログ回路に実装される [24] といった優れた面も多いが、ラインプロセスのしきい値が絶対的な値に固定されるのでグレイレベルのスケール変化で結果が変わる、セグメンテーションを得るにはエッジの補完など複雑な後処理を要する、インパルスノイズを除去するよう拡張もできるが求解が困難 [12]、画素の格子配列が仮定されランダムな配置の場合には適用できないなどの難点がある。そこでここではこれらの難点を緩和するため、ラインプロセスをある意味で相対化した変数である結合プロセスを導入した画像復元法を提案する。結合プロセスは心理学や生理学の分野で linking process [17] とか associative field [18] などと呼ばれているニューロン結合に対応するものである。

3.2 ラインプロセスによる画像復元

ラインプロセスを使う画像復元は次のような最適化問題として表される．ここでは，簡単のため1次元で記す．

$$\begin{aligned} \min \quad & \alpha \sum_i (d_i - f_i)^2 + \beta \sum_i (f_i - f_{i+1})^2 (1 - l_i) + \gamma \sum_i l_i \\ & + T \sum_i [l_i \ln l_i + (1 - l_i) \ln (1 - l_i)] \end{aligned} \quad (3.1)$$

ここで d_i は第 i 画素のグレイレベル， f_i はその復元値である．式 (3.1) からラインプロセス l_i は

$$l_i = \frac{1}{1 + e^{\frac{\gamma - \beta(f_i - f_{i+1})^2}{T}}} \quad (3.2)$$

と与えられ， $(f_i - f_{i+1})^2$ が γ/β より大きいと1になり小さいと0になる．従ってグレイレベルの不連続が $\sqrt{\gamma/\beta}$ より大きいエッジは保存され，それより小さなエッジはノイズとして平滑化される．このことはこれらのパラメータの値を変えることによりこのしきい値を変えられるという自由度がある反面，これらのパラメータ値の適切な設定に困難を生じる．また式 (3.1) ではインパルスノイズの両脇でラインプロセスが1になってインパルスノイズがそのまま保存される．この問題点への対処法はまずペナルティエネルギー $\sum \sum l_i l_{i+1}$ を (3.1) に付け加えることである．しかしそうするとインパルスノイズが平滑化され，その近傍のグレイレベルが大きく歪む．そこで更に (3.1) の第1項を

$$\alpha \sum_i (d_i - f_i)^2 (1 - m_i) \quad (3.3)$$

と変更して更に $\delta \sum m_i$ を付け加える．こうすれば高さが $\sqrt{\delta/\alpha}$ より大きいインパルスノイズが取り除かれる．しかしその結果パラメータの設定は更に難しくなり，エネルギーの非凸性も激しくなって最適解を求めるのが困難となる．またこの方法は画像の復元だけでなくセグメンテーションにも使われるが，複雑な後処理を必要とし，またドットパターンのようなランダムな配置の画素データのクラスタリングには適用できない．

3.3 結合プロセスによる画像の復元とセグメンテーション

そこで本章ではラインプロセスの以上の難点を解消するために式(3.1)を次のように変更する.

$$\begin{aligned}
 \min \quad & \alpha \sum_i (d_i - f_i)^2 + \sum_i \sum_{j \in W_i} x_{ij} [\beta \|a_i - a_j\|^2 + \gamma (f_i - f_j)^2] \\
 & + T \sum_i \sum_{j \in W_i} x_{ij} \ln x_{ij} \\
 \text{subj.to} \quad & \sum_{j \in W_i} x_{ij} = 1
 \end{aligned} \tag{3.4}$$

ここで W_i は第 i 画素の周りのウィンドウであり, i 自身は含まない. なお式(3.1)では i は1次元配列であったがここでは2次元配列の画素に適当に番号をつけたものである. a_i は第 i 画素の2次元座標値である. 変数 x_{ij} は式(3.1)のラインプロセス l_i に相当する変数であり, 第 i 画素と第 j 画素との近接度を表し, 両画素の場所とグレイレベルが近いほど1に近くなる. 式(3.4)のラグランジュ関数は

$$\begin{aligned}
 L = & \alpha \sum_i (d_i - f_i)^2 + \sum_i \sum_{j \in W_i} x_{ij} [\beta \|a_i - a_j\|^2 + \gamma (f_i - f_j)^2] \\
 & + T \sum_i \sum_{j \in W_i} x_{ij} \ln x_{ij} + \sum_i \lambda_i (\sum_{j \in W_i} x_{ij} - 1)
 \end{aligned} \tag{3.5}$$

となり, 最適解はこの関数の停留値である. まず

$$\begin{aligned}
 \frac{\partial L}{\partial x_{ij}} &= \beta \|a_i - a_j\|^2 + \gamma (f_i - f_j)^2 + T(\ln x_{ij} + 1) + \lambda_i \\
 &= 0
 \end{aligned} \tag{3.6}$$

と

$$\begin{aligned}
 \frac{\partial L}{\partial \lambda_i} &= \sum_{j \in W_i} x_{ij} - 1 \\
 &= 0
 \end{aligned} \tag{3.7}$$

から

$$x_{ij} = \frac{e^{-\frac{\beta \|a_i - a_j\|^2 + \gamma (f_i - f_j)^2}{T}}}{\sum_{k \in W_i} e^{-\frac{\beta \|a_i - a_k\|^2 + \gamma (f_i - f_k)^2}{T}}} \tag{3.8}$$

が求まり，これを L の f_i による勾配の式

$$\begin{aligned} \frac{df_i}{dt} &= -\frac{\partial L}{\partial f_i} \\ &= \alpha(d_i - f_i) + \gamma \sum_{j \in W_i} (x_{ij} + x_{ji})(f_j - f_i) \end{aligned} \quad (3.9)$$

に代入して平衡解を求めることにより復元値 f_i が求まる． l_i と x_{ij} の違いは式 (3.2) と式 (3.8) に現れている．ラインプロセスは隣り合う画素の間にしか存在せず，（このことはまた画素の隣接性が予め決められていることを前提としている）空間的な方向性はないが，結合プロセスはウインドウ内の全ての画素間に存在し，画素の空間的な関係は座標値によって表されており，空間的な方向性がある．すなわち x_{ij} は第 i 画素の第 j 画素への近接度を表し，一般に $x_{ij} \neq x_{ji}$ である．しかしもっと重要な違いは両者のグレイレベルの依存性である．すなわち l_i が f_i と f_{i+1} の絶対的な近接度で決まるのに対し， x_{ij} は相対的な近接度で値が決まる．すなわち各 i について位置及びグレイレベルが自分に最も近い j への x_{ij} が 1 に近くなる．この相対性は式 (3.4) で x_{ij} の総和を 1 に制約していることから生じている．式 (3.1) と (3.4) のエントロピー項の違いはこの制約条件からきている． x_{ij} と似た働きをするニューロン間の結合が生理学や心理学で扱われ，linking process[17] や associative field[18] などと呼ばれている．ここでは x_{ij} を結合プロセスと呼ぶ．式 (3.9) は x_{ij} あるいは x_{ji} で結合されたピクセルの間にだけグレイレベルの平滑化作用が働くことを表している．また一方 x_{ij} によりセグメンテーションが与えられる．ある画素がウインドウ内部の他の全ての画素と均等に結合したときの x_{ij} は，ウインドウ内部の他の画素の数を n とすると， $1/n$ である．そこで x_{ij} と x_{ji} のうちどちらか一つでも $1/n$ より大きければ両画素は結合しているとする．これによって結合している画素の集合がセグメントを形成するとする．このようにすればセグメントが直接与えられるのでラインプロセスのときのようにエッジを補完するような後処理は要しない．また， k 平均法などのクラスタリングアルゴリズムに基づく方法と違い，セグメントの数を予め決める必要もない．

なお本方法では T はある正の値に固定する．すなわちアニーリングは行なわない．これはパラメータの値によってはアニーリングによって得られる低エネルギーの解は必ずしも平滑性の高い解ではないことによる．同じことが 2 節のラインプロセスによ

る画像復元法でも指摘されている [25, 26]. この理由からここでは T は正の値に固定し, x_{ij} をファジイな値にすることによってパラメータ値の変化に対してロバストな結果が得られるようにした.

3.3.1 画像復元

式 (3.2) はラインプロセスのしきい値が絶対的であることを示し, 式 (3.8) は x_{ij} が相対的に決まることを示す. 従ってグレイレベルのスケール変換に対してラインプロセスよりも結合プロセスの方がより不変であると期待される. これを図 3.1 のような簡単なデータで調べた. これは階段状のデータにノイズを乗せたものである. このデータの縦のスケールすなわちグレイスケールを変化させて復元結果を調べた. ラインプロセスによる方法と結合プロセスによる方法とも定性的な振る舞いは同じで, スケールが小さいとジャンプエッジがなまり, スケールが大きいとノイズがエッジとして残る. エッジが保存されてノイズが平滑化される望ましい復元結果はこれらの上下のしきいスケールの間で得られる. ラインプロセスによる方法ではこれらのしきいスケールは 0.5 と 2.0 であったのに対し, 結合プロセスによる方法では 0.2 と 3.0 であった. このように結合プロセスによる方法の方がラインプロセスによる方法よりもグレイスケールの広い範囲で望ましい復元をすることができる.

本方法により平滑化を行った例を図 3.2 に示す. 図 3.2(a) は 256 階調の画像に ± 30 の範囲のガウスノイズを乗せたものであり, 図 (b) が復元結果である. パラメータの値は $\alpha = 0.0002$, $\beta = 0.01$, $\gamma = 0.01$, $T = 1.0$ である. ウィンドウは 9×9 とした.

3.3.2 インパルスノイズの除去

式 (3.4) ではインパルスノイズも平滑化されるのでその付近のグレイレベルが歪む. インパルスノイズは平滑化するのではなく取り除かねばならない. そこで式 (3.3) と同じような変更によってインパルスノイズの場所でエネルギーの第 1 項が 0 になるようにする. インパルスノイズは近傍のグレイレベルから離れているので, 第 i 画素がインパルスノイズとすると他の画素からの結合はないので $\sum_j x_{ji} \simeq 0$ となり, またこの

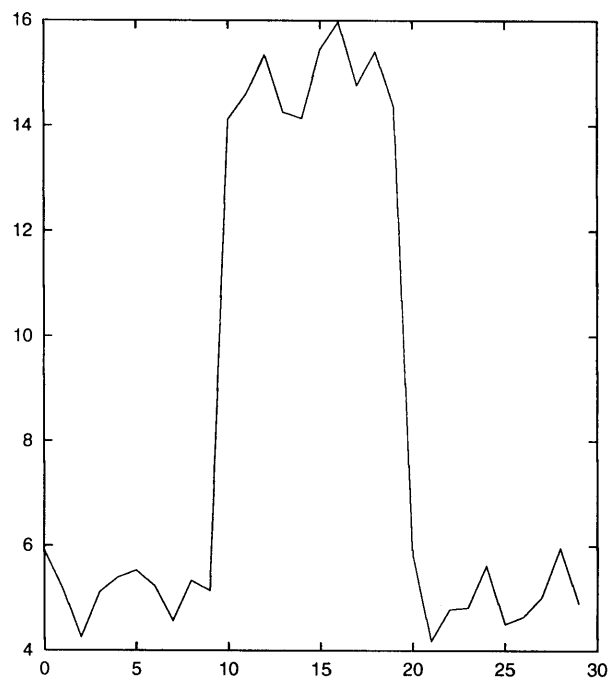
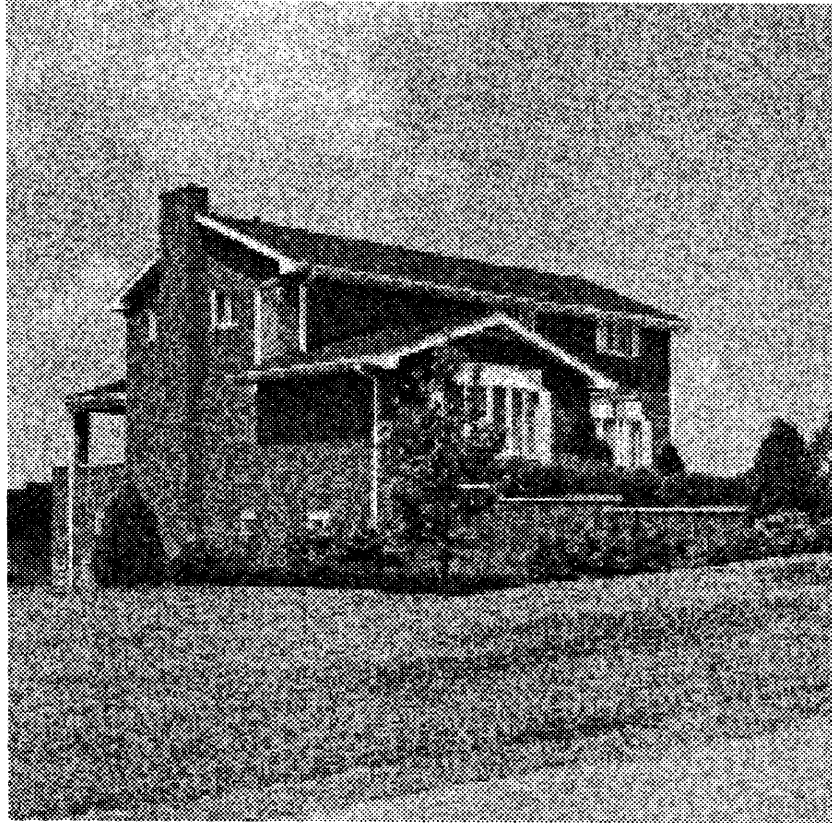


図 3.1: データ例.



(a)noisy image



(b)restored image

図 3.2: 画像復元の例.

ときに限り第 i 画素はインパルスノイズである。そこで式 (3.4) の第 1 項を

$$\alpha \sum_i \left(\sum_{j \in W_i} x_{ji} \right) (d_i - f_i)^2 \quad (3.10)$$

と変更すればよい。例として図 3.3(a) の階段データにガウスノイズとインパルスノイズを乗せた図 (b) のデータの復元結果を図 (c) に示す。比較のため図 (d) にメディアンフィルタの結果を示す。ウィンドウ幅は両方法とも 5 である。このようにガウスノイズの平滑化能力において本方法はメディアンフィルタよりも優れている。

3.3.3 セグメンテーション

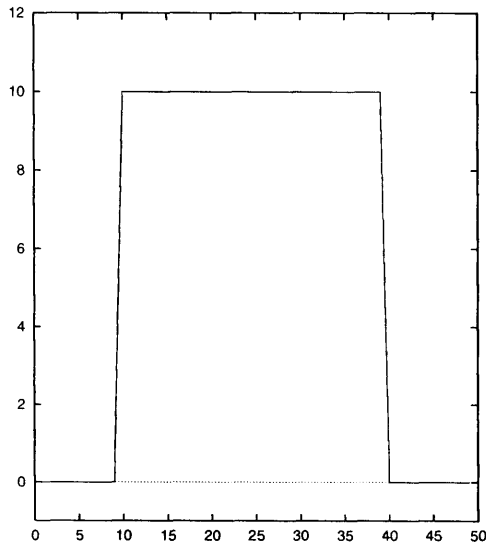
図 3.4 に図 3.2 の画像のセグメンテーションの結果を示す。これは結合プロセスで互いに結合している領域を同じグレイレベルで表示したものである。

3.4 マルチスケール表現

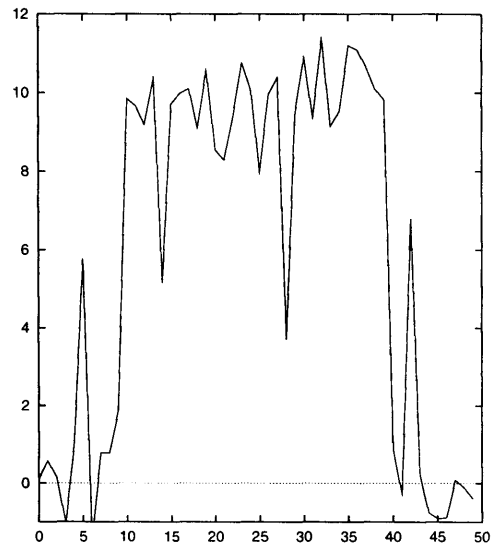
式 (3.1) でパラメータの一つを連続的に変化させるとエッジの保存性が連続的に変わりマルチスケール表現が得られる [25]。また同じようなマルチスケール表現は非等方拡散法という方法でも得られる [27]。本方法でも同様な操作により同様なマルチスケール表現が得られる。

3.4.1 連続変形によるマルチスケール表現

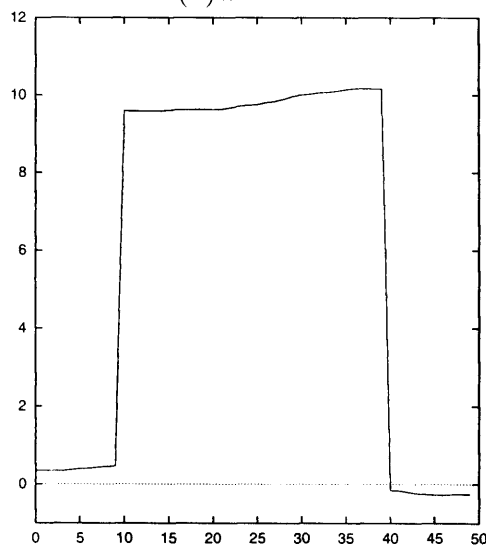
式 (3.4) で $\alpha \uparrow \infty$ とすると $f_i = d_i$ となり $\alpha \downarrow 0$ とすると全ての f_i が同じ値になる。従って α を連続的に小さくしていくとエッジがしだいになまっていく。このとき図 3.5 のように各エッジはある α の値で不連続になり、図 (a) のように突出部の幅が狭いエッジから、また図 (b) のように高さが低いエッジからそれぞれ先になまっていく。実際の画像ではこれらの変化が組み合わさって生じる。例として図 3.6 に 1 次元データ (脳の断層像のグレイレベル) のマルチスケール平滑の結果を示す。同図の x 軸に記した \times 印はセグメントの境界である。



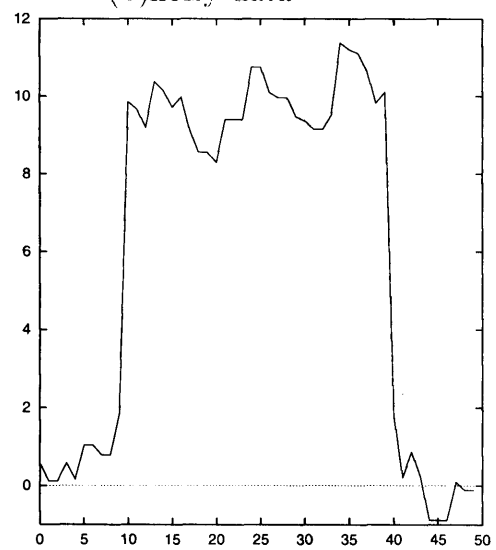
(a) data



(b) noisy data



(c) smoothed data



(d) median filtering

図 3.3: 1次元データの復元例.

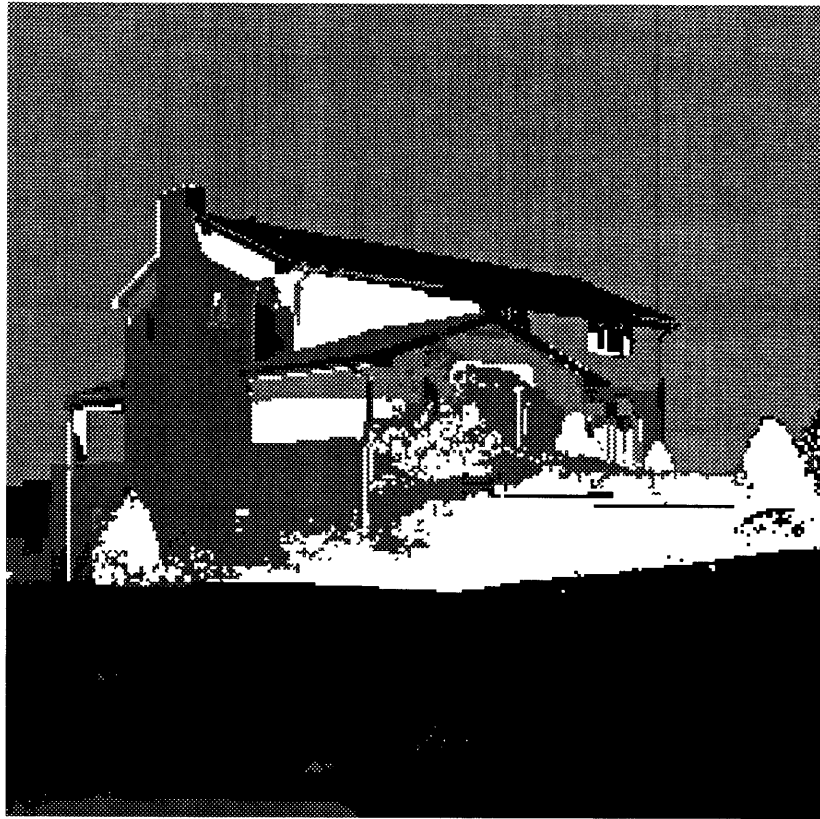
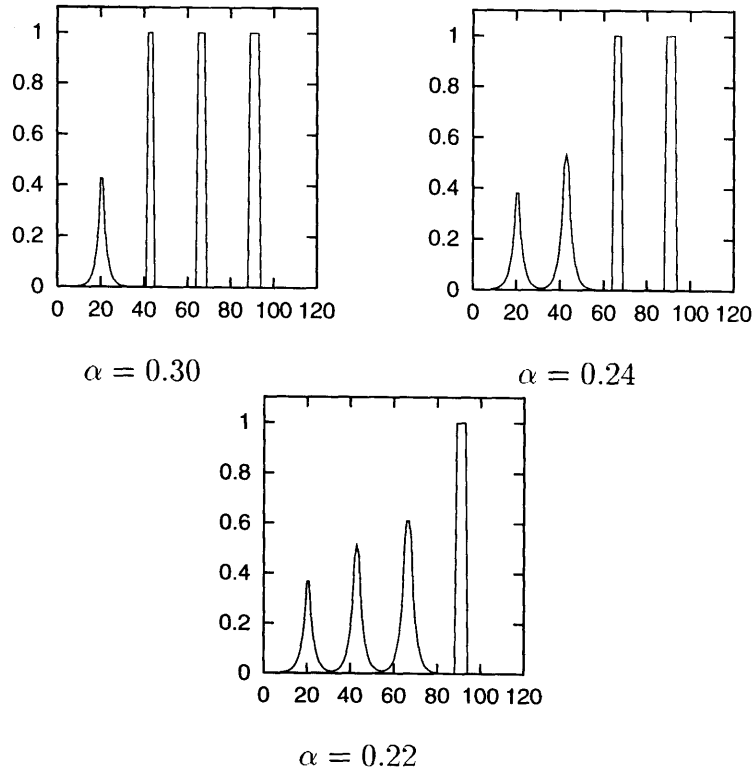
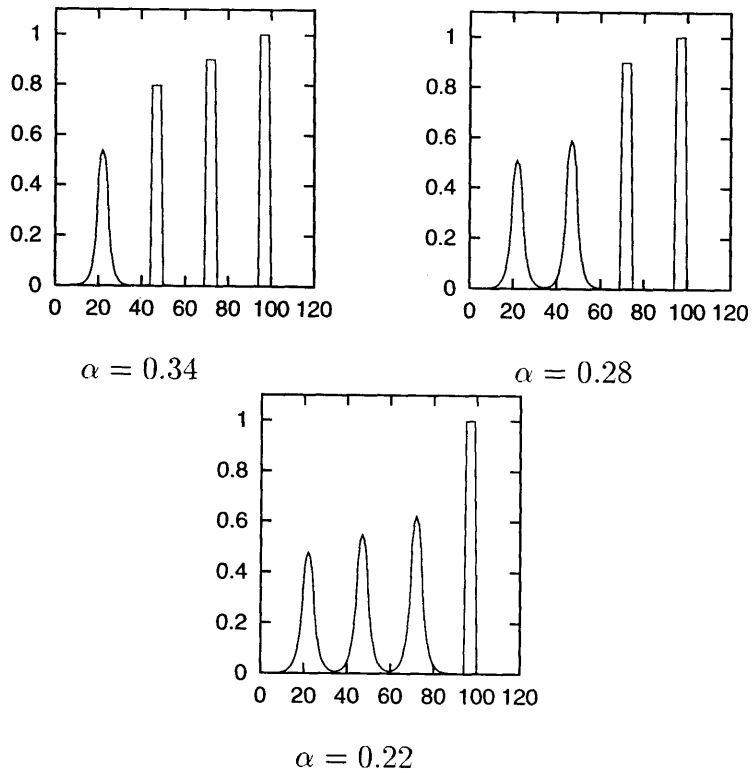


図 3.4: セグメンテーションの例.



(a)



(b)

図 3.5: α の減少によるエッジのなまり方.

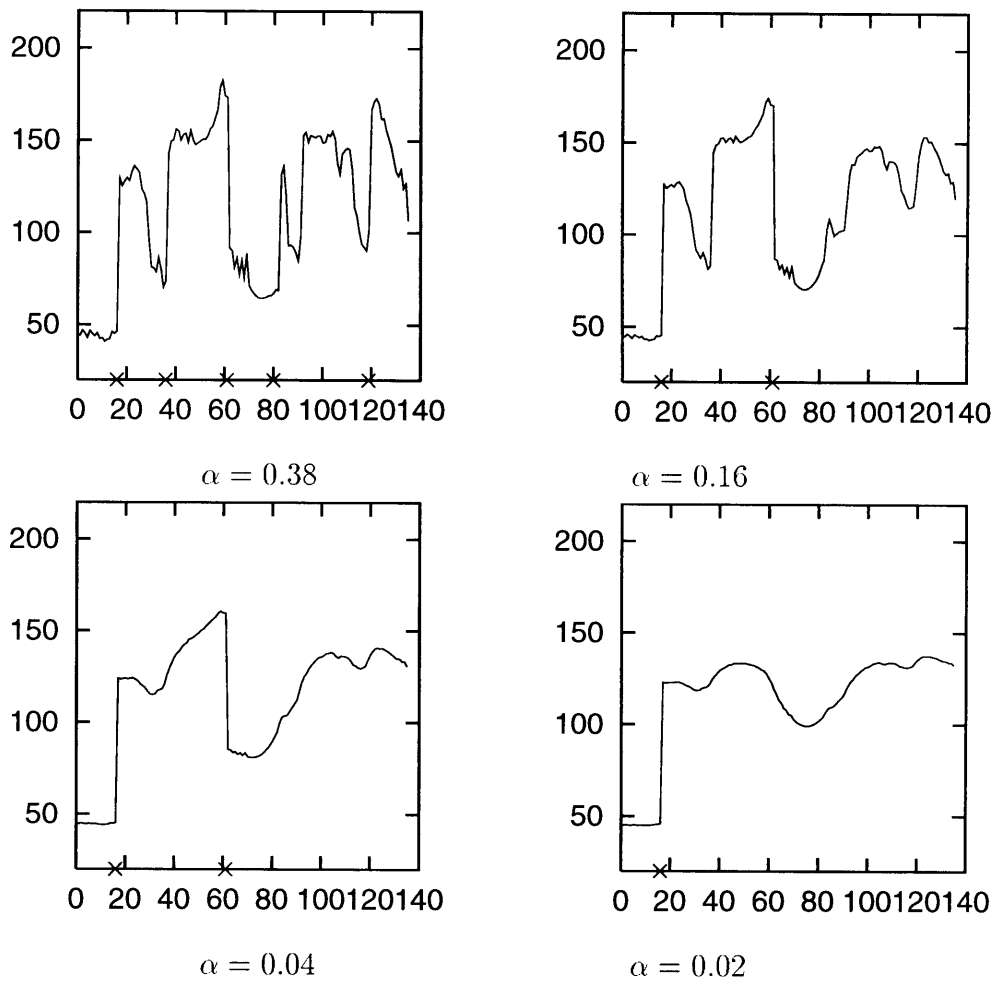


図 3.6: マルチスケール平滑化.

3.4.2 拡散によるマルチスケール復元

式 (3.9) において $\alpha = 0$ とおくと次式になる.

$$\frac{df_i}{dt} = \gamma \sum_{j \in W_i} (x_{ij} + x_{ji})(f_j - f_i) \quad (3.11)$$

この式は拡散係数が結合プロセスでコントロールされる非線形拡散方程式である. 式 (3.11) では $t \uparrow \infty$ の平衡解では全ての f_i が同じ値になる. そこで f_i の初期値を d_i として式 (3.11) の解を時間的に追っていけば前節とほとんど同じマルチスケール表現が得られる. 前節の方法ではパラメータの値を変えながら各値で式 (3.4) の解を求める必要があるが, この拡散による方法では単純に式 (3.11) の時間を追っていくだけでよく, 計算が格段に速い. 従ってマルチスケール表現を得るにはこの拡散による方法の方がよい. 図 3.7(a) に図 3.2 の画像のマルチスケール平滑の結果を, また図 (b) にそのときのセグメンテーションを示す.

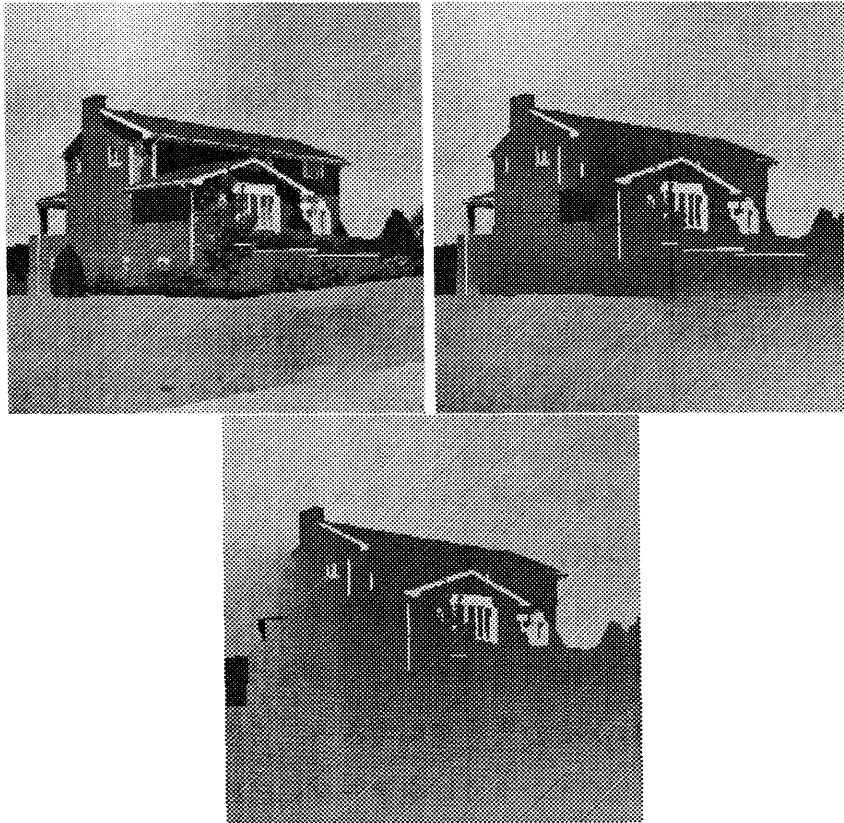
3.5 スパースなデータへの応用

測定系によってはスパースな画素でしかデータが得られないことがある. メディアンフィルタなどではこのような場合を取り扱うことができない. ラインプロセスの復元法ではこのようなスパースなデータからでも全画素でのグレイレベルを復元することができる. これは式 (3.1) の補間能力による. ここでは本方法でも同様な復元ができることを示す. 更に本方法ではスパースなデータそのものだけを扱う復元やセグメンテーションも行える. これはラインプロセスによる方法では扱えないものである.

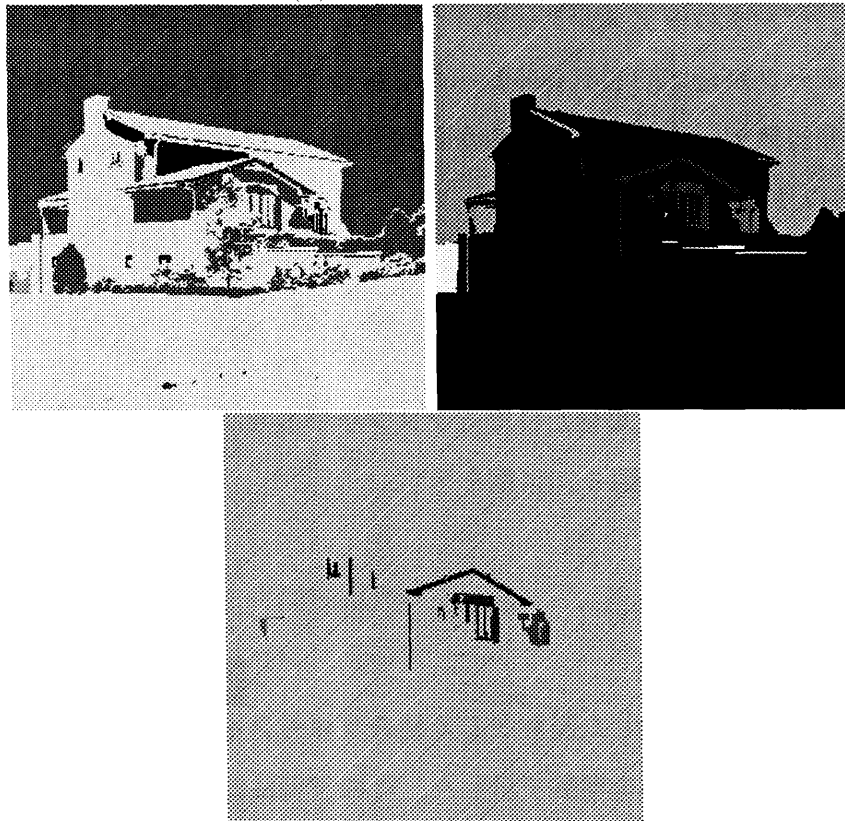
3.5.1 スパースデータからの全画像の復元

スパースにサンプリングされた画素でだけグレイレベルがデータとして与えられ, その他の画素ではデータが与えられない場合に, これらのスパースなデータから全画素でのグレイレベルを復元することは式 (3.4) の第 1 項を

$$\alpha \sum_i p_i (d_i - f_i)^2 \quad (3.12)$$



(a) smoothed image



(b) segmentation

図 3.7: マルチスケール平滑化とセグメンテーション.

と変更すれば行える。ここで p_i はデータが与えられる画素で1でその他の画素では0とする。図 3.8に原画像 (a) から 10% の割合でランダムに画素をサンプリングしたデータ (b) からの復元 (c) 例を示す。エッジに凹凸が生じているがデータがない画素へもグレイレベルが補間されている。

3.5.2 ドットパターンの復元とセグメンテーション

次に、与えられたスパースなデータだけを扱って、その他の画素への補間を行わない場合を考える。この場合は復元よりもセグメンテーションが主な目的である。ラインプロセスによる方法では、画素間の隣接性が予め与えられねばならず、ランダムな配置では隣接性を決めるのは困難であるから、このような状況を扱うのは難しい。それに対し本方法では式 (3.4) がそのまま適用できる。この場合画素の場所がスパースになっただけである。ただしウィンドウは全データに広げる。図 3.9に図 3.8のスパースデータに対する結果を示す。図 3.9(b) は x_{ij} で結合された画素同士を直線で結んだものであり、データのセグメンテーションを表している。更にもっと極端な場合としてスパースデータが全て同じグレイレベルを持つときを考える。この場合グレイレベルの復元値も当然データと同じになるから式 (3.4) で $f_i = d_i, f_i = f_j$ となる。従って x_{ij} は解析的に与えられ、

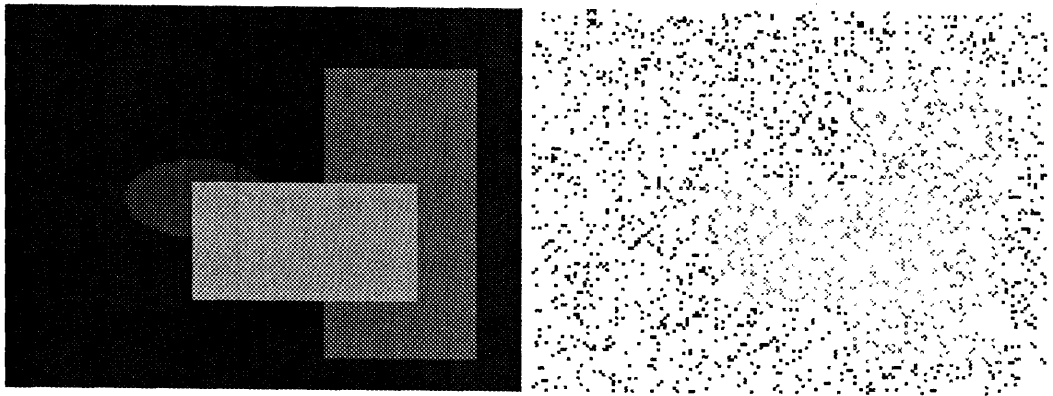
$$x_{ij} = \frac{e^{-\beta \frac{\|a_i - a_j\|^2}{T}}}{\sum_{k \in W_i} e^{-\beta \frac{\|a_i - a_k\|^2}{T}}} \quad (3.13)$$

となる。すなわち場所の近接性だけでセグメンテーションが与えられる。これはドットパターンの群化である。図 3.10に例を示す。

3.6 むすび

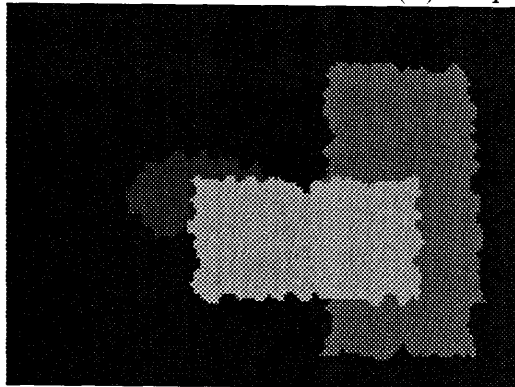
ラインプロセスを相対化した変数として結合プロセスを導入し、エッジを保存するノイズ平滑化ができることを実験で示した。この復元法においてパラメータを変化させたり、拡散方程式を用いたりするとマルチスケール復元ができる。またインパルス

ノイズも除去できるように拡張し，スパースなデータからの復元やセグメンテーションにも適用できることを示した．本方法がラインプロセスによる方法よりも優れている点は，グレイレベルのスケール変化の広い範囲で結果が変わらないこと，セグメンテーションが直接得られること，インパルスノイズの除去が容易に行えること，ドットパターンなどのスパースなデータのセグメンテーションにも適用できることなどである．一方ラインプロセスは隣接画素間にしかないのに対し結合プロセスはウインドウ中の全ての画素間にあるので，量が多く，計算時間がかかる．計算の高速化が今後の課題である．



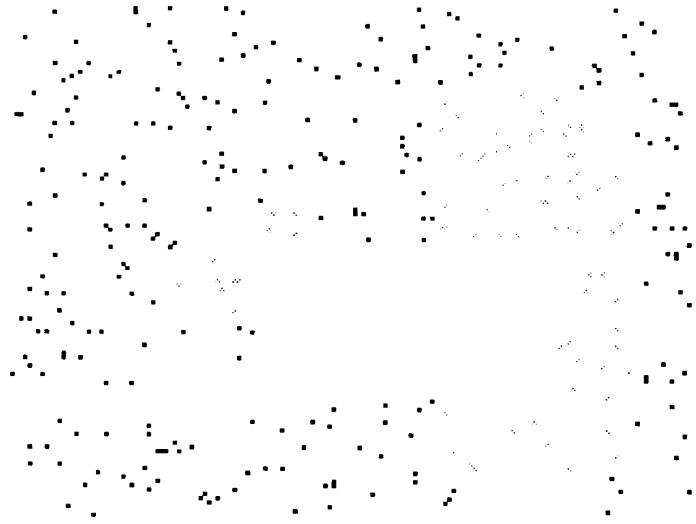
(a)image

(b)sampled data

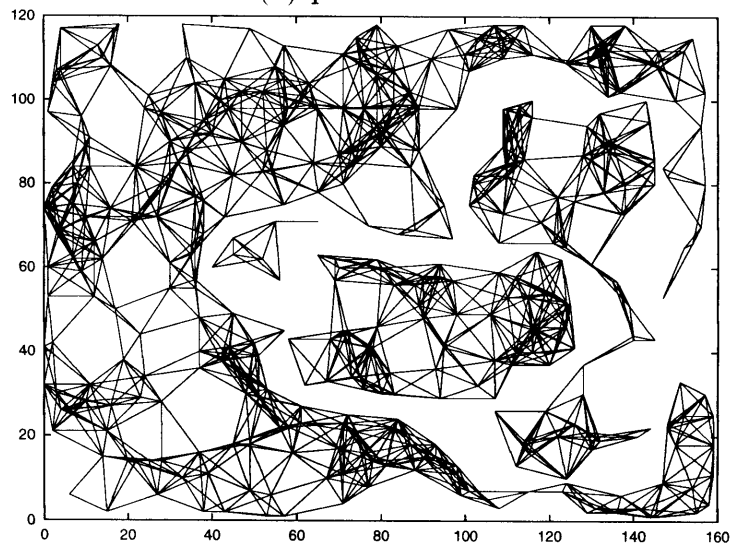


(c)restored image

図 3.8: スパースなデータからの復元.

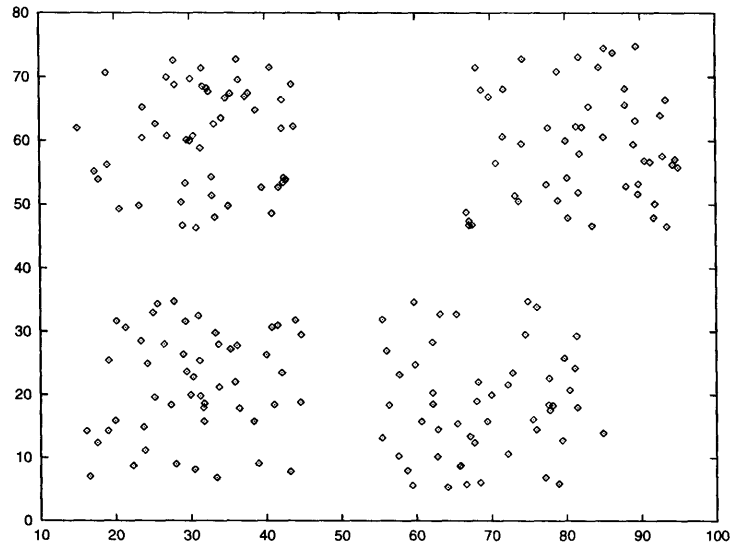


(a)sparse data

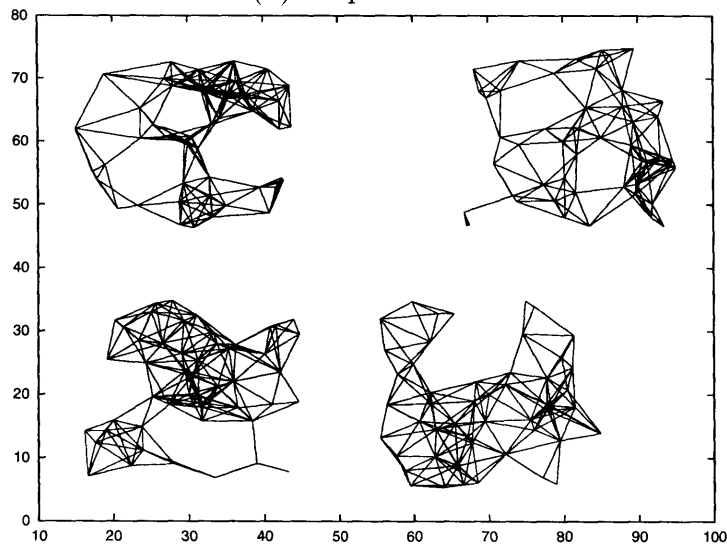


(b)linking

図 3.9: スパースデータのセグメンテーション.



(a)dot pattern



(b)linking

図 3.10: ドットパターンのクラスタリング.

第 4 章

結合プロセスによるドットパターンの群化モデル

4.1 まえがき

人間にはパターンを見たときそれを体制化する強い傾向があり，それは計算の高速化に役立っていると Watt[28] は推測している．パターンの体制化のモデルはこれまで数多く提案されているが，本章では新しい単純なモデルを提案する．ゲシュタルト心理学におけるプレグナンツの法則のように体制化の要因には沢山あるので，ここではこの要因が最も少ないようにドットパターンに限定する．ドットはそれ自身の特徴としては位置だけしか持たない．従ってドットの体制化で最も強く働くのは近接要因であると考えられる．交差する直線上に並ぶドットでの交点付近など近接性以外の要因が働く場合もあるが，ここでは最も基本的なモデルとして近接要因だけによる群化のモデルを提案し，そのモデルによってどの程度の心理現象を再現できるか調べる．ここで提案するモデルはクラスタリングのモデルであるが，これは副次過程としてドットの変位を伴っており，これによって錯視のようなパターンの変形が説明できるという立場をとる．本モデルを作るに当たり大前提とする心理実験結果の性質は非線形性とスケール不変性である．非線形性とはパターンの体制化にヒステリシス現象が観察されることを指す．これを大前提とすると心理学での場の理論 [29] や同様な網膜の受

容野に基づくモデル [30] や重ね合わせに基づく線形モデル [31] あるいはグラフ理論によるクラスタリング [32] などはヒステリシスを生じず知覚されるパターンは一意に決まるので、これらのモデルは排除される。そこでここでは協調系と呼ばれるクラスに含まれるモデルを提案する。またスケール不変性はここではパターンが拡大、縮小されてもクラスタリング結果は不変であり、更にドットの変位量はパターンの拡大率に比例する、すなわち相対的な変位量はスケール不変であることを表す。これは錯視の心理実験で実証されている [33]。またパターンの体制化は何らかの最適化処理の結果であると考え、ここでは最適化問題としてモデルを定式化する。

4.2 結合プロセスモデル

ドットの実際の位置を d_1, \dots, d_n とし、それらのドットはこの真の位置からずれた位置 r_1, \dots, r_n にあるように知覚されるとする。各 d_i や r_i は 2 次元座標 $d_i = (d_{ix}, d_{iy})$, $r_i = (r_{ix}, r_{iy})$ である。ここでは r_i は次の最適化問題の解として得られるというモデルを提案する。

$$\begin{aligned}
 \min \quad & \alpha \sum_{i=1}^n \|d_i - r_i\|^2 + \beta \sum_{i=1}^n \sum_{j=1, \neq i}^n x_{ij} \|r_i - r_j\|^2 + T \sum_{i=1}^n \sum_{j=1, \neq i}^n x_{ij} \ln x_{ij} \\
 \text{subj.to} \quad & \sum_{j=1, \neq i}^n x_{ij} = 1 \quad (i = 1, \dots, n) \\
 & x_{ij} \geq 0 \quad (i, j = 1, \dots, n)
 \end{aligned} \tag{4.1}$$

ここで x_{ij} は前章で述べた結合プロセスであり、第 i ドットと第 j ドットとの間の結合の強さを表す。 α, β, T は正定数である。エネルギーの最初の項は元々の位置 d_i への復元力を表し、第 2 項はドット間の引力を表す。第 3 項はエントロピーであり、 x_{ij} を 0-1 整数でなく $[0, 1]$ 間のファジイな値にする効果を持つ。温度 T が 0 だと x_{ij} は 0 か 1 となり、各ドットは自分に最も近いドットとだけ結合し、他のドットとは全然結合しないが、 T が大きくなるにつれて最近接以外のドットへの結合も生じてきて T が無限大では距離によらず全てのドットと同じ割合 $x_{ij} = 1/(n-1)$ で結合する。後述するように心理実験との比較から、 T の値は適度な正の値であると推測される。式 (4.1)

の最適解はラグランジュ関数

$$L = \alpha \sum_{i=1}^n \|d_i - r_i\|^2 + \beta \sum_{i=1}^n \sum_{j=1, \neq i}^n x_{ij} \|r_i - r_j\|^2 + T \sum_{i=1}^n \sum_{j=1, \neq i}^n x_{ij} \ln x_{ij} + \sum_{i=1}^n \lambda_i \left(\sum_{j=1, \neq i}^n x_{ij} - 1 \right) \quad (4.2)$$

の停留点

$$\frac{\partial L}{\partial r_i} = 2\alpha(r_i - d_i) + 2\beta \sum_{j=1, \neq i}^n (x_{ij} + x_{ji})(r_i - r_j) = 0 \quad (4.3)$$

$$\frac{\partial L}{\partial x_{ij}} = \beta \|r_i - r_j\|^2 + T(\ln x_{ij} + 1) + \lambda_i = 0 \quad (4.4)$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{j=1, \neq i}^n x_{ij} - 1 = 0 \quad (4.5)$$

である。(4.4)と(4.5)からラグランジュ乗数 λ_i を消去すると x_{ij} が

$$x_{ij} = \frac{e^{-\beta \|r_i - r_j\|^2 / T}}{\sum_{k=1, \neq i}^n e^{-\beta \|r_i - r_k\|^2 / T}} \quad (4.6)$$

と得られる。従ってあとは残る(4.3)の解を微分方程式

$$\frac{dr_i}{dt} = \alpha(d_i - r_i) + \beta \sum_{j=1, \neq i}^n (x_{ij} + x_{ji})(r_j - r_i) \quad (4.7)$$

の平衡解として求めればよい。(4.6)を(4.7)に代入すれば(4.7)は r の非線形な微分方程式となり、引力の係数 $(x_{ij} + x_{ji})$ は(4.6)から分かるように r_i と r_j との距離の非線形な関数であるので、(4.7)は一般に多安定である。すなわちある与えられた d で式(4.7)は複数個の安定な平衡解を持ち、 r がどの平衡解に落ち着くかは r の初期値によって異なる。この r の初期値については次のようにとることにする。通常の場合、すなわちパターンを十分な時間間隔をおいて見せられるときには r の初期値は d とする。特殊な場合、例えばよく似たパターンを短い時間間隔で引き続いて見せられた場合には r の初期値は直前のパターンでの r の平衡値にとる。初期値をこのようにとることによってヒステリシス現象を再現できる。

次にこのモデルのスケール不変性について考えよう。ここでのスケール不変とは r が d に比例する、すなわち r/d が d によらず一定であることを意味する。まず最初に

検討すべきことは式(4.1)のエネルギーでユークリッド距離の2乗を採用していることである。これが2乗以外だと式(4.3)の $r_i - d_i$ や $r_j - r_i$ は1乗ではなくなりスケール不変ではなくなる。従ってこれは2乗でなければならない。また x_{ij} もスケール不変でなければならない。温度 T が非常に小さいときには x_{ij} は0か1となり、最も近い r_i と r_j の x_{ij} だけが1になる。これはスケールが変わっても変わらない。従って T が十分小さいときにはこのモデルはスケール不変である。しかし後述の心理実験結果との比較から推測されるように T は適度に正の値であることが強く推測される。これは一般的に実際の知覚がファジイであることやニューロンの特性がステップ関数でなくなめらかなシグモイド関数であることなどからも妥当なことであると思われる。しかし式(4.6)の形から、 T が大きいと x_{ij} はスケール不変ではなくなる可能性が推測され、実際シミュレーションしてみるとスケール不変ではなかった。そこでここではスケールによって温度 T が変化するという仮定を設定することにする。すなわち T は $\max \|d_i - d_j\|^2$ に比例すると仮定する。このようにすると $\|r_i - r_j\|^2/T$ はスケール不変になり式(4.6)から x_{ij} がスケール不変になる。すなわちファジイ度がスケールによらず一定になる。逆に言えば、ファジイ度がスケールによって変わるのは不自然であるからこの仮定は妥当であろう。

4.3 シミュレーション結果

このモデルの第一義の目的はドットパターンのクラスタリングであり、変数 x_{ij} がクラスタ分割を決める。このクラスタ分割が副次過程としてドットの変位を伴っているという設定であり、この変位によっていくつかの錯視が再現できるであろうとする立場である。そこでまず基本的な目的の処理であるクラスタリングについてシミュレーションしてみる。パラメータは $\alpha = 1$, $\beta = 0.2$, T のスケールの比例係数 = 0.05とした。

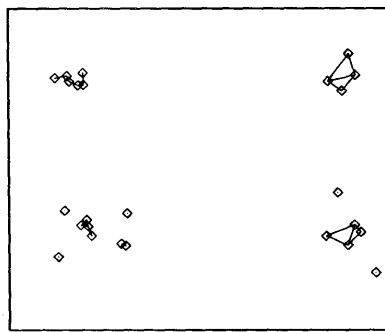
4.3.1 ドットパターンの群化

ドットパターン d_1, \dots, d_n が与えられたとき上記のモデルで x_{ij} を求めると各ドット間の相対的な近接度が分かる。第 i ドット d_i と第 j ドット d_j の結合度は x_{ij} と x_{ji} で与えられる。そこでこれらの両方が、あるしきい値以上であるとき、すなわち $\min(x_{ij}, x_{ji}) > \varepsilon$ であるときこれら二つのドットを線で結合することにし、そのようにして作った無向グラフの連結成分をクラスタとすることにする。このドット間の連結線は Stevens[34] の仮想線に相当するものである。このようにすると α , β , T あるいは ε といったパラメータを変えると階層的なクラスタリングが得られる。なかでも最も簡単なのはしきい値 ε を変える場合であり、 x_{ij} を 1 回計算すればその結果を使って ε のいろいろな値でグラフを作れば階層的なクラスタリングが得られる。図 4.1 にそのようにして得られたグラフの例を示す。

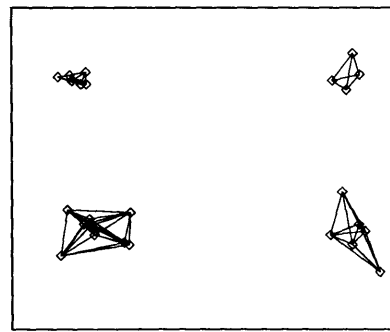
次に本モデルを非線形にした理由であるヒステリシスについてシミュレーションしてみた。図 4.2 のドット (\diamond 印) を上の列から 1 列ずつ順次見せていくと真ん中の列 (ここではドットは全て等間隔で並んでいる) まで来てもそれまでと同じクラスタリングが知覚される。見せる順序を逆にして下の列から見せると同様なことが起こり、ヒステリシスがあることが分かる。このヒステリシスはシミュレーションによって図 4.2 に示しているように再現された。 \diamond 印がドットの本来的位置 d 、 \times 印が知覚される位置 r である。しきい値 ε をほぼ 0 から 1 近くまで変えても図 4.2 は変わらない。このシミュレーションでは各列での r の初期値は直前の列での r の平衡値とした。

4.3.2 錯視

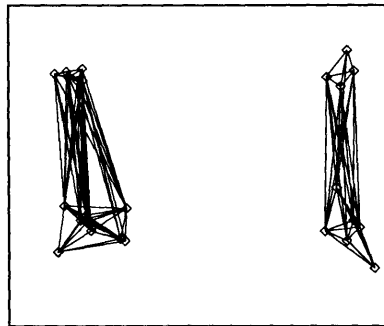
本モデルの本来的な目的であるクラスタリングにおいてはドットの d から r への変位は副次的なものでしかないが、知覚過程で実際にこのような変位が生じていると主張する立場をここではとる。この立場が正当であるかどうかについてはそのような変位が顕著に知覚されるいわゆる錯視について検証してみればよいと思われる。そこで多くの錯視図形は線で構成されているがそれを等間隔のドットで置き換えてシミュレーションしてみた。線をドットで置き換えて近接性だけを扱くと、線が本来持っていた



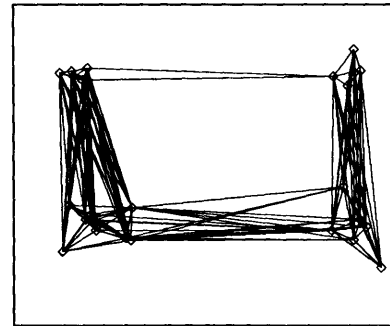
(a) $\varepsilon = 0.2$



(b) $\varepsilon = 0.01$



(c) $\varepsilon = 1 \times 10^{-12}$



(d) $\varepsilon = 1 \times 10^{-19}$

図 4.1: ドットのクラスタリング

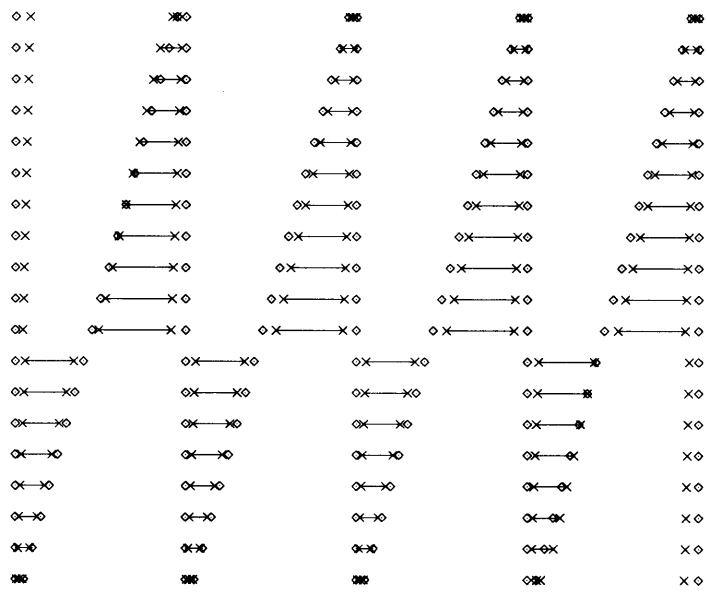


図 4.2: ドット結合のヒステリシス

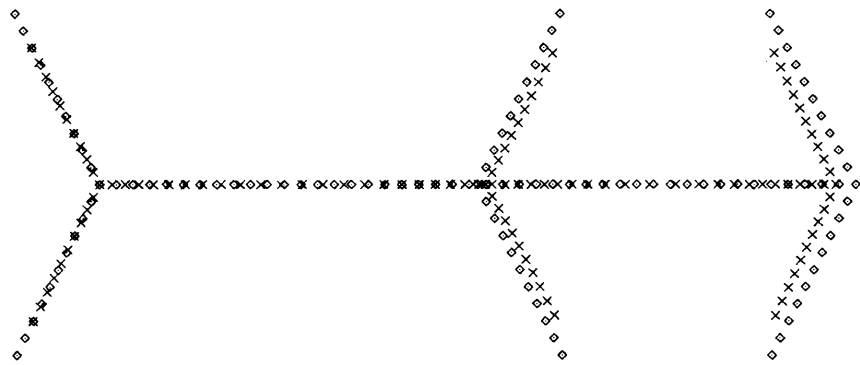


图 4.3: Muller-Lyer 图形

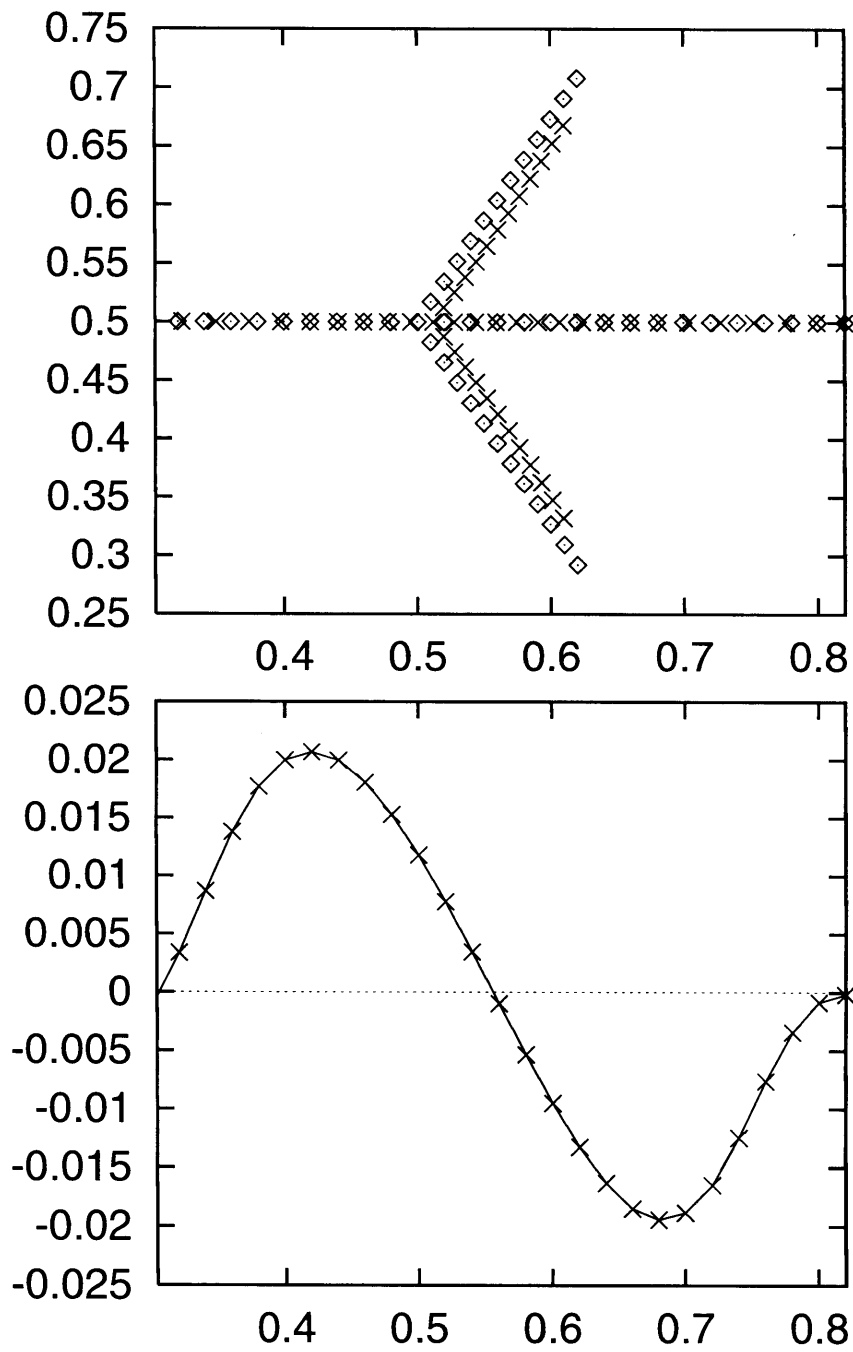


図 4.4: ドットの変位のシミュレーション結果

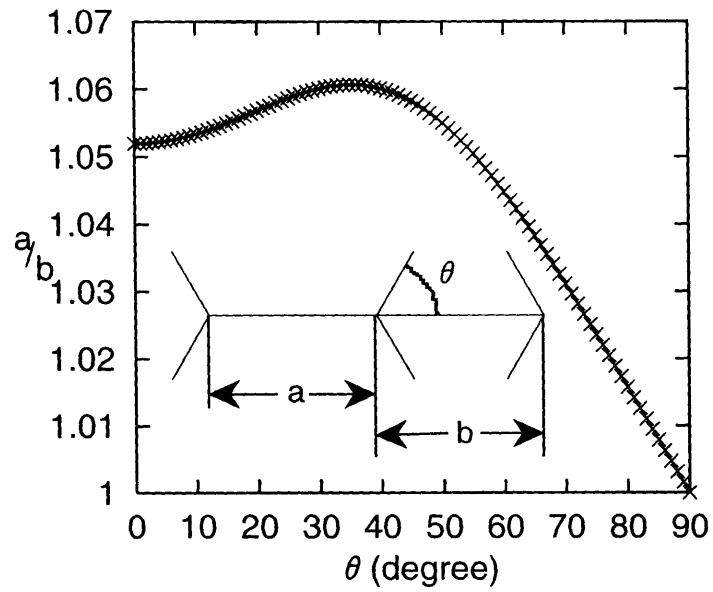


図 4.5: 矢線の角度による錯視量の変化

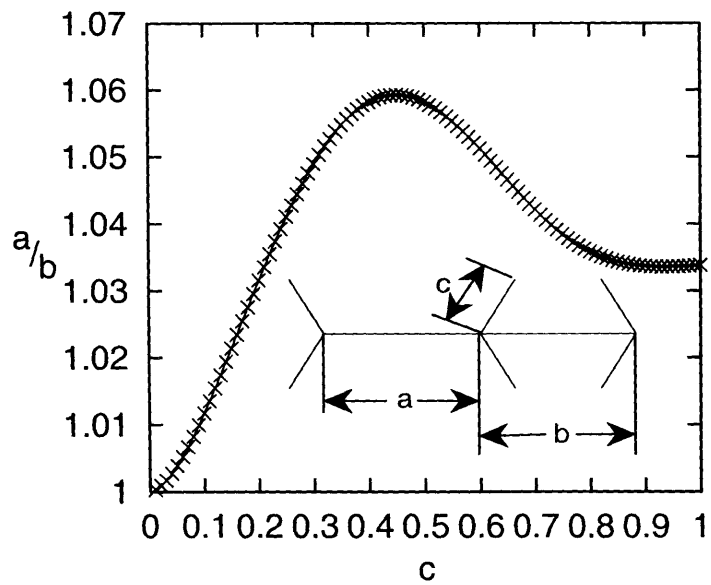


図 4.6: 矢線の長さによる錯視量の変化

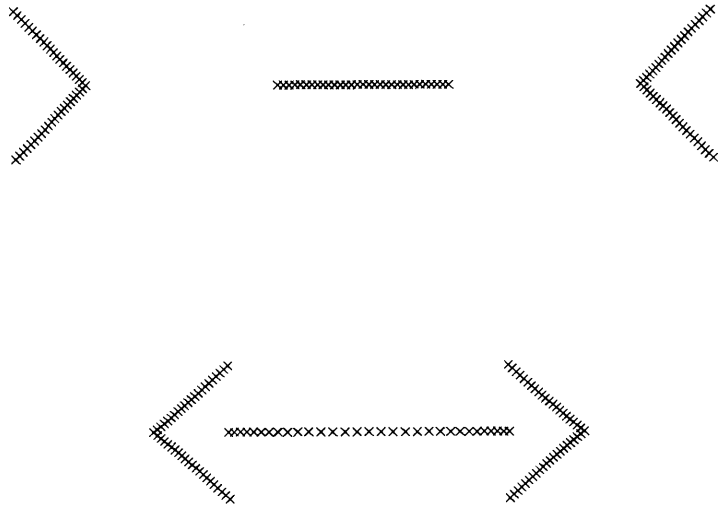


图 4.7: 逆錯視图形

向きの情報が落ちてしまうがなおそれでも再現できる錯視現象もあるはずである。逆に言えばドットで置き換えても再現できるような錯視の主要因は近接性であるといえる。まず最も有名な Muller-Lyer 図形について図 4.3 のように錯視が再現できた。もう少し半定量的に検討するために中央の矢線の交点付近での主線上のドットの変位を調べた。図 4.4 にシミュレーション結果をしめす。変位量の形および不動点の位置ともよく心理実験 [33] を再現できている。また図 4.5 は矢線の角度による錯視量の変化、図 4.6 は矢線の長さによる変化であり、両方とも変化の形および最大になる角度や長さとも心理実験結果 [33] とほぼ一致している。また心理実験では図 4.7 のように図形を変形した場合本来の錯視とは逆の錯視すなわち図 4.7 の上の主線は下の主線よりも短く見えるといういわゆる逆錯視が観察されるがこれについても図 4.7 のように再現できた。この図には r のみをプロットしてある。元々の主線は上下同じ長さである。以上のように Muller-Lyer 図形は近接要因だけでもある程度半定量的に再現できた。

次に同心円錯視についてシミュレーションしてみた。中心が同じ二つの円を書き、内側の円を固定して外側の円の半径を増やしていくと内円は過大視され、その錯視量は内円と外円の半径比が 2 : 3 で最大となりそれ以後減少し、外円の半径が十分大きくなると逆に内円の過小視が起こる [33]。この図で横軸は内外円の半径比の対数であり、縦軸は相対的な錯視量である。外円を固定して内円を次第に小さくした場合には外円の過小視のみが起こり、その錯視量はやはり内外半径比 2 : 3 で最大になる [33]。これらの図で線がそれぞれ 3 本あるのは図の大きさを 3 通り変えた実験結果であり、これらがほとんど重なっていることは錯視量が絶対距離でなく相対距離であることすなわちスケール不変であることを示す。これに関する本モデルのシミュレーション結果を図 4.8 に示す。このシミュレーションでも図の大きさを 3 通り変えてみたがそれらの結果は全て完全に重なった。内円の錯視では過大視から過小視に移ることや外円の錯視では過小視しか起こらないことなど心理実験とよく似た結果が得られた。なお以上のシミュレーションにおいては温度 T は適度の正の値としている。すなわちクラスタリングは適度にファジイである。 T が非常に小さい値だとクリスプなクラスタリングとなり、例えば図 4.8(a) は図 4.9 のようになる。これはドット間の距離があるしきい値より大きい小さいかでドット間の引力がステップ関数状に変化することによ

る。同様に T が小さいと図 4.4, 図 4.5, 図 4.6 の曲線もある点で不連続に変化するようになる。このことから逆に推測すると視覚でのクラスタリングはある程度ファジイであると思われる。

次に図 4.10 の分割距離錯視図形について中央の点 (すなわち右から 2 番目の点) と左端の点との間の点の数を変化させてシミュレーションしてみたところ, 中央の点から左端の点までの距離と右端の点までの距離はそれぞれ図 4.11 のように変化した。左半分が右半分よりも長く見えることおよび左半分の距離が点の数が増えるにつれて山型に変化することなど心理実験 [33] と定性的に一致している。なおクラスタリングとしては左半分が一つのクラスタにまとまり, 右端の点は常時切り離されている。

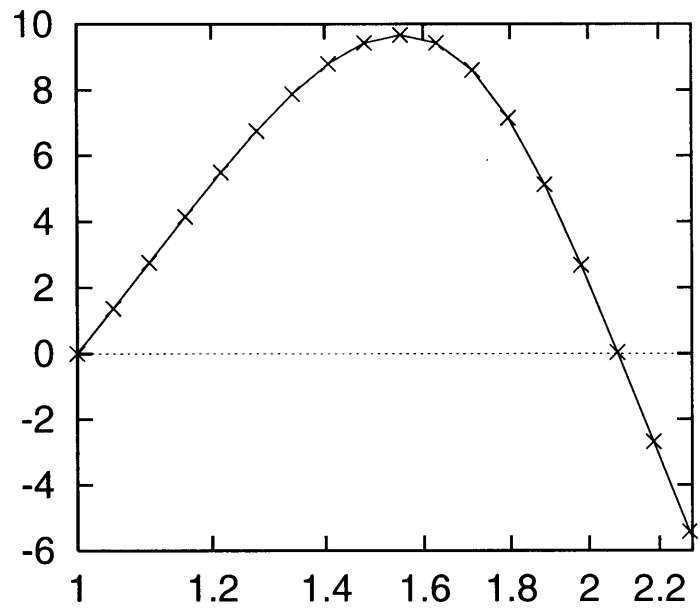
次に曲率の錯視についてシミュレーションした。円弧や半円形は縮んで見え, 短い円弧ほど曲率が小さく知覚される。図 4.12 にシミュレーション結果を曲率を添えて示す。元の円弧の曲率はすべて 1 である。

他の錯視図形で本モデルである程度再現できた例のシミュレーション結果を図 4.13 から図 4.20 に示す。これらの図ではすべて r だけをプロットしてある。元の図形 d では周知のように, 図 4.13 では左右の内円は同じ大きさ, 図 4.14 では左の外円と右の内円が同じ大きさ, 図 4.15 では垂直の線と水平の線は同じ長さ, 図 4.16 では正方形, 図 4.17 では 3 本の線の中央部分は平行, 図 4.18 では上の斜めの線の延長は下の線の右端と交わる, 図 4.19 では上と下の平行線は同じ幅, 図 4.20 では 7 個の円の下端は水平である。これらのなかには線の向きに関する錯視がいくつか含まれているが, 線の向きに関する他の図形, 例えばツェルナー, ポッケンドルフ, ヘリング, ブント図形などについては本モデルでは錯視を再現できなかった。

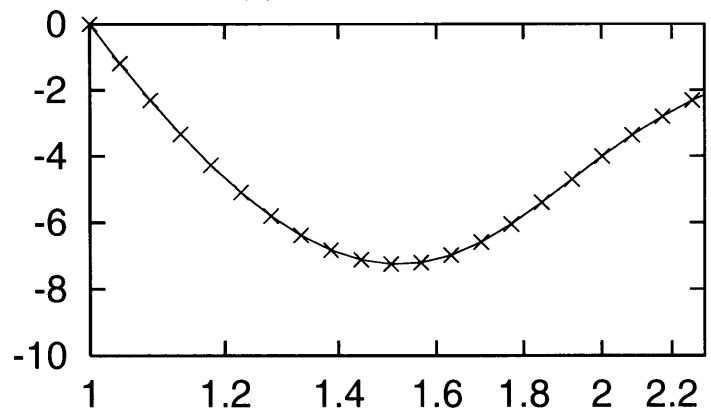
4.4 むすび

ドットパターンのクラスタリングの簡単な非線形モデルを提案し, 近接性のみによるクラスタリングやある種の錯視が再現できることを確認した。ここで再現できたものよりも高次の群化や錯視を再現するためにはエネルギーに高次の項, 例えば線の向きに関する項などを付け加えなければならないであろう。またそこまで行かなくても

例えばドットに大きさや形などの特徴がある場合には式(4.1)の質量係数 α をドットごとに変えて α_i とし, 引力係数 β もドットペアごとに β_{ij} としなければならない. 更に本モデルはスケールに対してもまた回転に対しても不変であるが, 実際には錯視は垂直方向と水平方向で大きく違い, パターンの回転に関して異方性を示す. このように本モデルについて更に拡張, 修正すべき点は多い.



(a) 内円の錯視



(b) 外円の錯視

図 4.8: 同心円錯視のシミュレーション結果

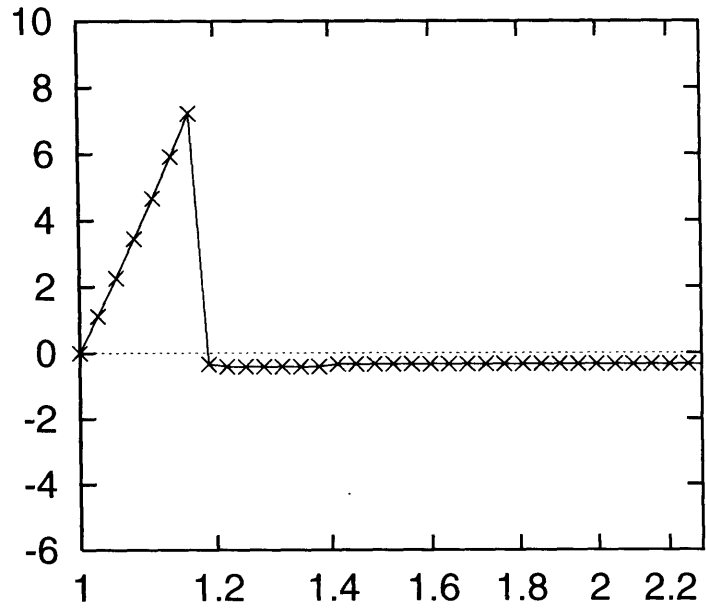


図 4.9: $T = 0$ のときの図 10(a)

◇ ◇ ◇ ◇ ◇ ◇ ◇ ◇ ◇ ◇

図 4.10: 分割距離錯視図形

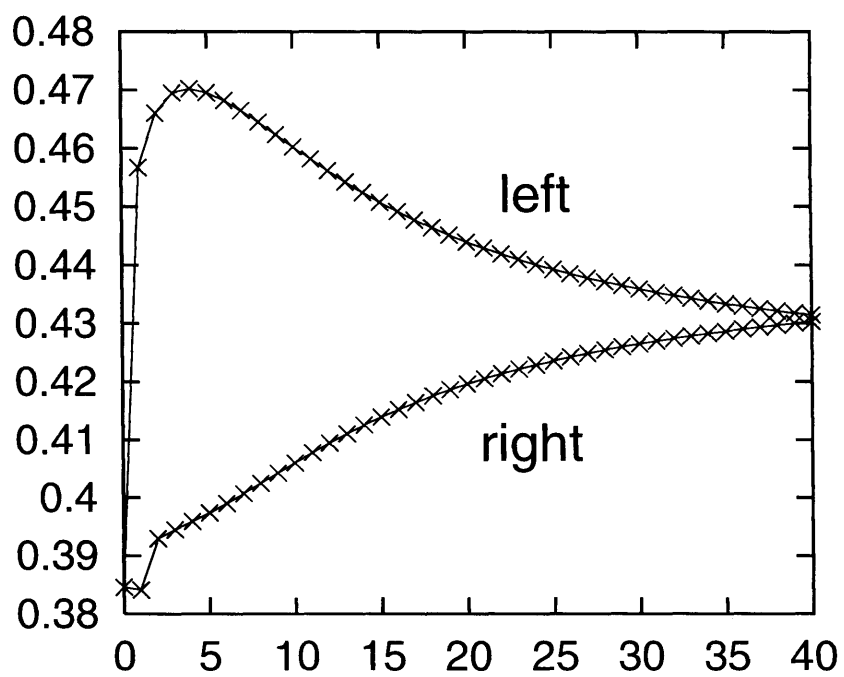


図 4.11: ドット数による左右の長さの変化

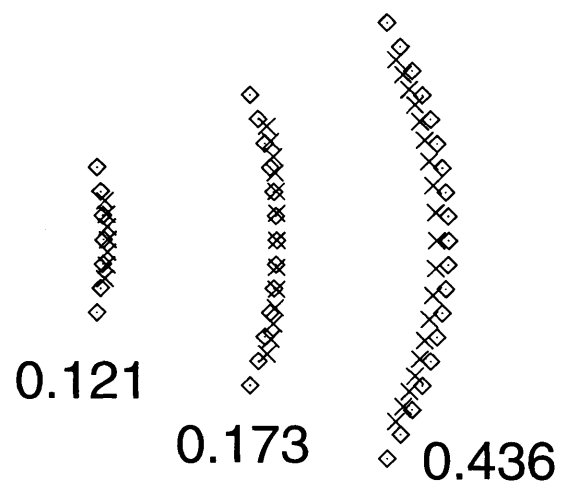


図 4.12: 曲率の錯視

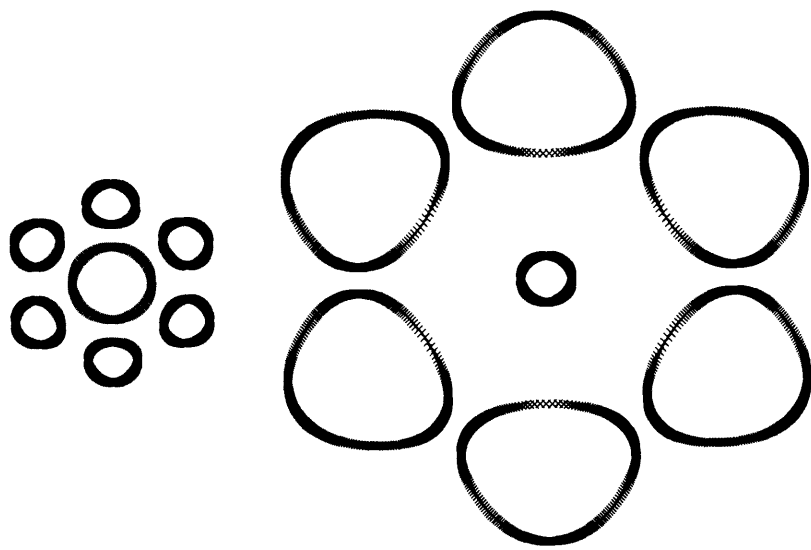


图 4.13: Ebbinghaus 图形

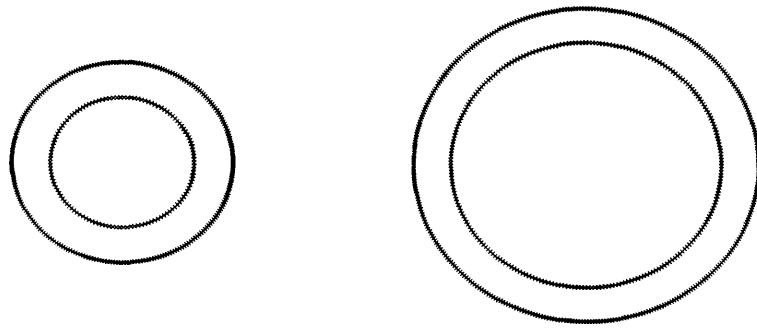


图 4.14: Delboeuf 图形

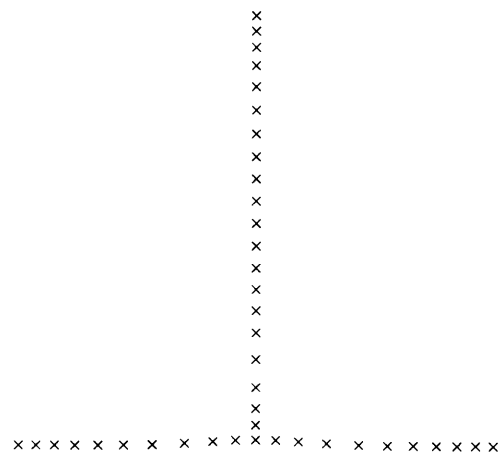


图 4.15: Fick 图形

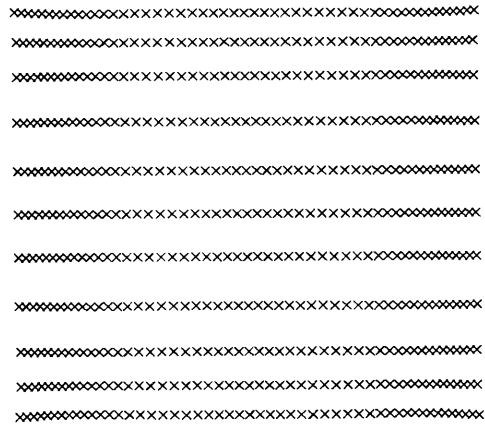


图 4.16: Helmholtz 四角形

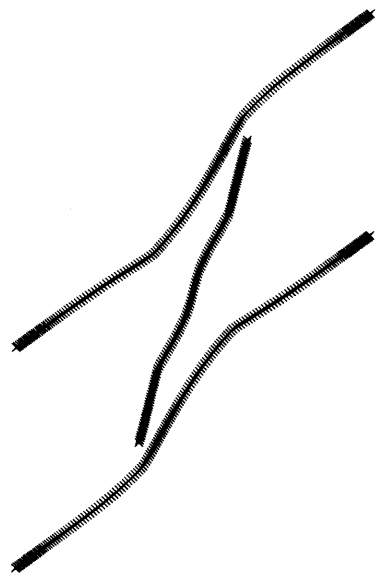


图 4.17: Lipps 图形

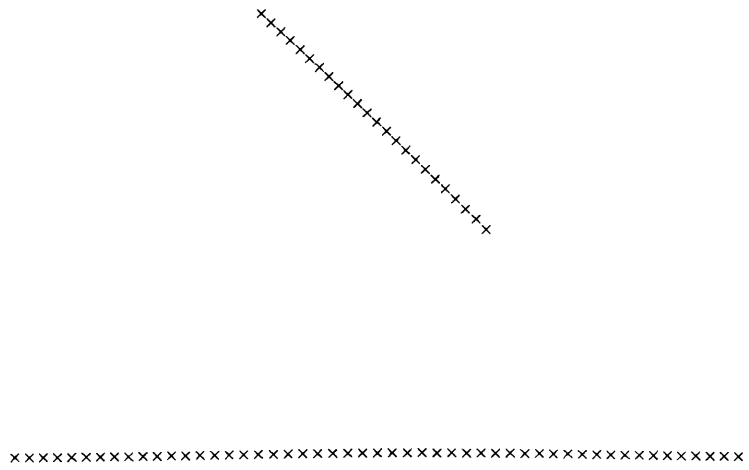


图 4.18: 角度方向錯視图形

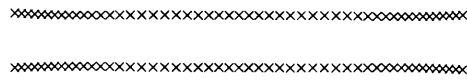
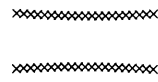


图 4.19: 平行線幅錯視图形

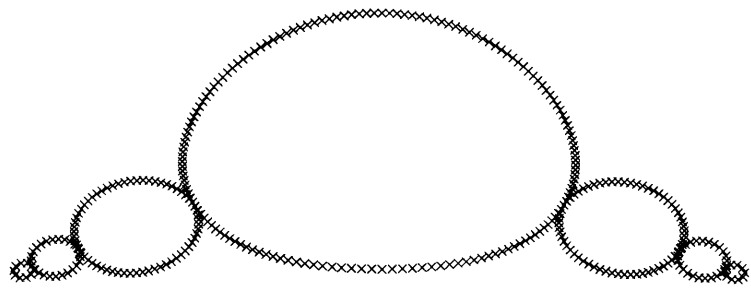


图 4.20: 湾曲错视图形

第 5 章

クラスタリング回帰による最近傍識別器 の学習法

5.1 まえがき

最も単純な最近傍識別器では訓練データ全部のなかからテストデータに最も近いものが選択される。しかしこれでは記憶量や探索時間が膨大となるので、各カテゴリーに適当な数の代表点が置かれ、それらの中からテストデータに最も近いものを選択する方がよい。このとき代表点の最適な配置が問題となる。ニューラルネットでは学習によって代表点を最適な配置に動かしていく。最も代表的な学習法は Kohonen[35] の Learning Vector Quantization (LVQ) である。この LVQ は勾配系ではないが、勾配系に修正した方法も提案されている [62]。またファジイ最近傍識別に基づく学習法も提案されている [58, 63]。しかしこれらの学習法には、各カテゴリーの代表点の数を予め設定しなければならず、また木探索識別器には適用できないなどといった欠点がある。そこでここではこれらの難点を解消する方法として関数回帰に基づく学習法を提案する。本学習法は、識別器の学習が関数が定数であるという関数回帰の特別な場合であることに基づいて、独立変数の空間をクラスタ分割するような関数回帰の学習法 [64] によって最近傍識別器の代表点の最適な配置を求めるものである。本方法の特長は、十分な数の代表点を用意しておけば各カテゴリーに必要な十分な数の代表点が自

動的に割り振られることと、木探索識別器の代表点を一括して最適化する学習が行えることである。簡単な例題によってこれらの性質を検証する。

5.2 最近傍識別器

パターン識別器は、テストパターンに対しそれが各クラスに属すメンバーシップ値を出力する。メンバーシップ値 $m(d)$ は 0,1 の整数値をとるので、テストパターン d の不連続な関数となる。ここでは最近傍識別器の学習法として勾配法に基づく方法 [58] を考えるので、連続な関数の極限としてメンバーシップ関数を表す。代表点を $r_j (j = 1, \dots, n)$ とするとテストデータ d の各代表点 r_j へのメンバーシップ x_j は

$$x_j(d) = \frac{e^{-\|d-r_j\|^2/T}}{\sum_{k=1}^n e^{-\|d-r_k\|^2/T}} \quad (5.1)$$

の $T \downarrow 0$ の極限值として与えられる。代表点 r_j の各クラス k へのメンバーシップ y_{jk} は既知であるので、 d の各クラス k へのメンバーシップ $m_k(d)$ は $y_{jk}x_j$ で与えられる。識別器の出力はこのメンバーシップそのものでよいが、一般的には各クラスで異なる値をとる任意の関数の値でよい。そうすると識別器の学習は関数近似として定式化され、識別関数の近似誤差を最小にする代表点 r_j を求めることになる。

5.2.1 教師なし学習

学習データとして $d_i (i = 1, \dots, m)$ が与えられるとする。近似される関数を恒等写像 $f(d) = d$ とし、これを区分的定数関数 $\sum_j x_j(d_i)r_j$ で近似することを考える。 $x_j(d_i)$ は式 (5.1) の $T \downarrow 0$ の極限值である。このとき近似誤差の最小化問題は次のように定式化される。

$$\min_{r_j} \sum_{i=1}^m \left\| d_i - \sum_{j=1}^n x_j(d_i)r_j \right\|^2 \quad (5.2)$$

$x_j(d_i)$ は 0 か 1 であるので、式 (5.2) は

$$\min_{r_j} \sum_{i=1}^m \sum_{j=1}^n x_j(d_i) \|d_i - r_j\|^2 \quad (5.3)$$

と同値である。これはクラスタリングすなわちベクトル量子化である。ここでは式(5.3)の解を勾配法で求めるために、式(5.3)に式(5.1)を代入して T も r_j と合わせて最適化することにする。すなわち

$$\frac{dr_j}{dt} = - \sum_{i=1}^m x_j(d_i)(r_j - d_i) - \sum_{i=1}^m \sum_{l=1}^n \frac{\partial x_l(d_i)}{\partial r_j} \|d_i - r_l\|^2 \quad (5.4)$$

$$\frac{dT}{dt} = - \sum_{i=1}^m \sum_{j=1}^n \frac{\partial x_j(d_i)}{\partial T} \|d_i - r_j\|^2 \quad (5.5)$$

によって最適解を求める。 T の初期値は適当に大きな値にする。 r_j はどのような初期値から出発しても、 T が大きいと一旦全データの重心に集まるので、初期値をデータの重心付近にランダムにとる。本学習法が通常の(決定論的)アニーリングと異なる点は、アニーリングでは T を変数でなくパラメータとして強制的に0まで減少させるが、ここでは式(5.5)によって変化させることである。このようにしても T が0に収束することは次のようにして分かる。まず式(5.3)の目的関数は $x_j(d_i)$ に関して1次であるから $x_j(d_i)$ の最適解は許容領域の端点となる。式(5.1)から分かるように $x_j(d_i)$ の許容領域は $[0, 1]$ であるから最適解では $x_j(d_i)$ は0か1になる。 d_i も r_j も有界であるから $x_j(d_i)$ が0か1になるのは $T \downarrow 0$ のときだけである。従って最適解では T は0に収束する。

5.3 クラスタ分割関数回帰

m 個のデータ $(d_i, f_i); i = 1, \dots, m$ に関数を回帰する問題を考える。 d_i, f_i ともベクトルであり、 d_i が独立変数で f_i は関数値である。ここでは独立変数の空間を n 個の領域に分割して、各領域で $g_j(d, \theta_j); j = 1, \dots, n$ という関数を当てはめる区分的関数による回帰をすることにする。 θ_j はパラメータである。この区分的関数は

$$f(d) = \sum_{j=1}^n x_j(d)g_j(d, \theta_j) \quad (5.6)$$

と表される。 x_j は各領域の特性関数すなわち第 j 領域で $x_j(d) = 1$ となり、その他の領域で $x_j(d) = 0$ となる関数である。式(1)の関数をデータ f_i に最小2乗フィッティ

ングする問題は

$$\min_{x_j, \theta_j} \sum_{i=1}^m \|f_i - f(d_i)\|^2 \quad (5.7)$$

となる。すなわち関数の2乗誤差の総和を最小にする分割 p とパラメータ r を求めることになる。特性関数 x_j の値は0か1であり、かつ任意の x で $\sum_j x_j(d) = 1$ であるから

$$\|f_i - f(d_i)\|^2 = \sum_{j=1}^n x_j(d_i) \|f_i - g_j(d_i, \theta_j)\|^2 \quad (5.8)$$

となり、従って式(5.7)は

$$\min_{x_j, \theta_j} \sum_{i=1}^m \sum_{j=1}^n x_j(d_i) \|f_i - g_j(d_i, \theta_j)\|^2 \quad (5.9)$$

となる。この式は、各データ f_i は自分に最もよくフィットする関数 g_j に割り当てられ(すなわち $\|f_i - g_j(d_i, \theta_j)\|$ が最も小さい j の x_j が1となり)、また各 g_j は自分に割り当てられたデータ f_i に最小2乗フィッティングされる(すなわち $x_j(d_i)$ が1になる i についてだけ g_j からの誤差が総和され最小化される)ということを表す。

5.3.1 クラスタリング回帰

x の分割法として単純なクラスタリングすなわちポロノイ分割を使うことにする。すなわち x が g_j に割り当てられるメンバーシップ $x_j(d)$ を式(5.1)の $T \downarrow 0$ の極限值とする。式(5.1)の値は $T > 0$ では0と1の間のファジイな値となるが、 $T \downarrow 0$ では0か1のクリスプな値となる。このようにすると式(5.9)は

$$\min_{r_j, T, \theta_j} \sum_{i=1}^m \sum_{j=1}^n x_j(d_i) \|f_i - g(d_i, \theta_j)\|^2 \quad (5.10)$$

となる。ただし x_j には式(5.1)を代入する。式(5.10)では T も最適化されることに注意されたい。式(5.10)の目的関数は $x_j(d_i)$ に関して1次であるから $x_j(d_i)$ の最適解は許容領域の端点となる。式(5.1)から分かるように $x_j(d_i)$ の許容領域は $[0, 1]$ であるから最適解では $x_j(d_i)$ は0か1になる。 d_i も r_j も有界であるから $x_j(d_i)$ が0か1になるのは $T \downarrow 0$ のときだけである。従って最適解では T は0になる。式(5.10)の最適解

は勾配降下系：

$$\begin{aligned}\frac{dr_j}{dt} &= -\frac{\partial E}{\partial r_j} \\ &= -\sum_{i=1}^m \sum_{k=1}^n \|f_i - g_k(d_i, \theta_k)\|^2 \frac{\partial x_k(d_i)}{\partial \theta_j}\end{aligned}\quad (5.11)$$

$$\begin{aligned}\frac{dT}{dt} &= -\frac{\partial E}{\partial T} \\ &= -\sum_{i=1}^m \sum_{j=1}^n \|f_i - g_j(d_i, \theta_j)\|^2 \frac{\partial x_j(d_i)}{\partial T}\end{aligned}\quad (5.12)$$

$$\begin{aligned}\frac{d\theta_j}{dt} &= -\frac{\partial E}{\partial \theta_j} \\ &= \sum_{i=1}^m x_j(d_i) [f_i - g_j(d_i, \theta_j)] \frac{\partial g_j}{\partial \theta_j}\end{aligned}\quad (5.13)$$

が収束した値として求まる。 E は式 (5.10) の目的関数である。式 (5.11) や (5.12) の右端の偏微分は式 (5.1) から容易に導かれるのでその表式はここでは省略する。最初 T は適度に大きな値にする。勾配降下系は局所最適解の一つに収束するから T は 0 に収束する。すなわち最終的にクリスプな分割が得られる。これは決定論的アニーリングであり、これによって大域最適解に近い解が得られる。すなわち式 (5.1) は勾配降下法が使えることを可能にし ($T \downarrow 0$ では式 (5.1) は微分できない) , かつアニーリングも自動的に行われる。

5.3.2 階層クラスタリング回帰

クラスタリングを木探索に拡張する。このときの学習法として1段目から順次後段へと学習をしていく逐次学習法と全ての段をまとめて学習する一括学習法とがある。逐次学習法は前節の学習を1段目から順次していくだけである、すなわちまず1段目の学習により x がボロノイ分割されるから、次の2段目では各ボロノイ領域について同様なクラスタリング回帰を行えばよい。従って逐次学習の説明は省略して一括学習について考える。簡単のため2段で説明する。 x は木探索ボロノイ分割される。1段目の分割の代表点を r_j , 2段目の代表点すなわち r_j の子を r_{jk} とする。 x の r_{jk} への

メンバーシップは

$$x_{jk}(d) = \frac{e^{-x_j(d)} \|d - r_{jk}\|^{2/T_2}}{\sum_l e^{-x_j(d)} \|d - r_{jl}\|^{2/T_2}} \quad (5.14)$$

と表される．ただし $x_j(d)$ には式 (5.1) で T を T_1 とおいたものを代入する．式 (5.14) は，親の r_j に属す (すなわち $x_j(d) = 1$ である) データだけを更に細分割したものが x_{jk} であることを表す．この細分割の各領域での関数を $g_{jk}(d, r_{jk})$ とすると回帰は次の最適化問題となる．

$$\min_{r_j, r_{jk}, \theta_{jk}, T_1, T_2} \sum_{i,j,k} x_j(d_i) x_{jk}(d_i) \|f_i - g_{jk}(d_i, \theta_{jk})\|^2 \quad (5.15)$$

ここで x_j は式 (5.14) のなかと同じものすなわち式 (5.1) で T を T_1 とおいたものである．代表点は r_j, r_{jk} と 2 段分あるが関数は g_{jk} しかないことに注意されたい．逐次学習では 1 段目で g_j が必要である．式 (5.15) の勾配降下系は

$$\frac{dr_j}{dt} = - \sum_{i,l,k} \frac{\partial x_l(d_i)}{\partial r_j} x_{lk}(d_i) \|f_i - g_{lk}(d_i, \theta_{lk})\|^2 \quad (5.16)$$

$$\frac{dr_{jk}}{dt} = - \sum_{i,l} x_j(d) \frac{\partial x_{jl}(d_i)}{\partial r_{jk}} \|f_i - g_{jl}(d_i, \theta_{jl})\|^2 \quad (5.17)$$

$$\frac{d\theta_{jk}}{dt} = \sum_i x_j(d_i) x_{jk}(d_i) [f_i - g_{jk}(d_i, \theta_{jk})] \frac{\partial g_{jk}}{\partial \theta_{jk}} \quad (5.18)$$

$$\frac{dT_1}{dt} = - \sum_{i,j,k} \frac{\partial x_j(d_i)}{\partial T_1} x_{jk}(d_i) \|f_i - g_{jk}(d_i, \theta_{jk})\|^2 \quad (5.19)$$

$$\frac{dT_2}{dt} = - \sum_{i,j,k} x_j(d_i) \frac{\partial x_{kj}(d_i)}{\partial T_2} \|f_i - g_{jk}(d_i, \theta_{jk})\|^2 \quad (5.20)$$

となる． T_1 と T_2 の収束については式 (5.15) の目的関数が多重線形すなわち x_j と x_{jk} の各々について 1 次であるから T_1, T_2 とともに 0 に収束する．

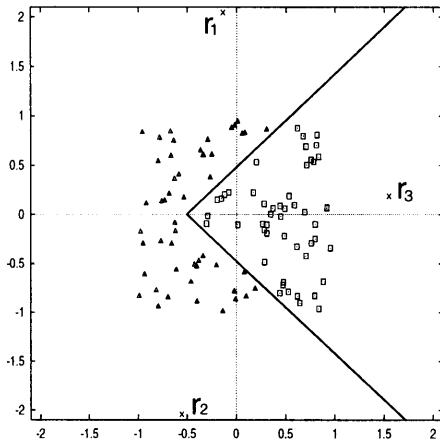
5.4 最近傍識別器の教師あり学習

式 (5.1) は最近傍識別器そのものである．従って上記のクラスタリング学習は最近傍識別器の教師あり学習にそのまま使える．最近傍識別器の学習では各学習データがどのカテゴリーに属すかが与えられる．そこである任意の次元の関数を考え，その関

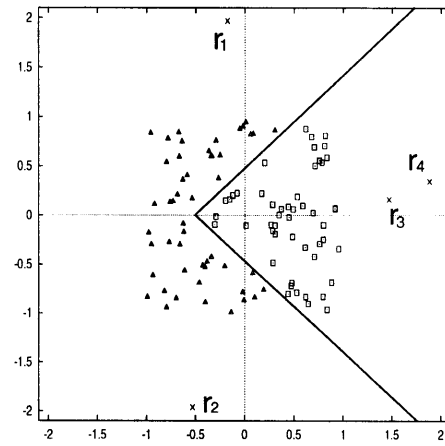
数が各カテゴリーでとる値も任意に与えれば，最近傍識別器の学習は上記の関数回帰の特殊な場合すなわち区分的定数データへの回帰に帰着される．関数の次数も任意でよいが，データが定数であるので定数関数で十分である．この関数は1種の識別関数である．ただし最近傍識別器では学習した後では関数は不要であり，各カテゴリーの代表点だけによって最近傍識別を行うことができる．学習法についてはクラスタリング学習において関数が定数という最も簡単な特例であるので省略する．このような最近傍識別器の学習で疑問になるのが各カテゴリーへの代表点の割り当て方と識別関数の次数の設定である．代表点の割り当てについては本方法で予め設定するのは代表点の総数だけであることに注意されたい．学習の結果それらが各カテゴリーに自動的に分配される．

5.4.1 代表点の割り当て

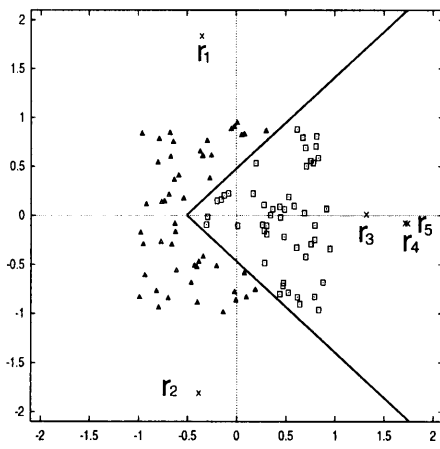
L V Qなどでは各カテゴリーの代表点の数は適当に（通常は十分多く，そして各カテゴリーに同数ずつ）設定される．しかし識別の計算速度や汎化などの観点からはカテゴリーそれぞれで必要最小限の数の代表点を使うのがよいと思われる．しかしそのようにカテゴリーごとに異なる代表点の最適な個数を予め知るのは困難である．そこで上記のクラスタリング学習で代表点が各カテゴリーにどのように分配されるかを簡単な例で調べてみた． x は2次元とし，カテゴリーは二つとする．従って識別関数は1次元とし，一方のカテゴリーで値0，他方のカテゴリーで値1をとるとした．まず図5.1のような100個のデータでやってみた．第1カテゴリーのデータを三角で，もう一方のカテゴリーのデータを四角で示してある．この例では代表点は左のカテゴリーに2個，右のカテゴリーに1個で必要十分である．代表点の総数を3個，4個，5個，6個としたときの結果はそれぞれ図5.1の(a)，(b)，(c)，(d)のようになった．図中の折れ線は代表点による識別境界であり，正しく識別されている．図(a),(b),(c),(d)とも r_1, r_2, r_3 が識別に使われ，図(b)では r_4 が，(c)では r_4 と r_5 が，また(d)では r_4, r_5 と r_6 が余り，これらの余剰代表点にはどのデータも所属しなかった．どの r_j が余るかは初期値による．代表点の初期値はデータの重心付近にランダムに配置したが， T が最初大きな値であるので初期配置がかなりばらついていても一旦データの中心付



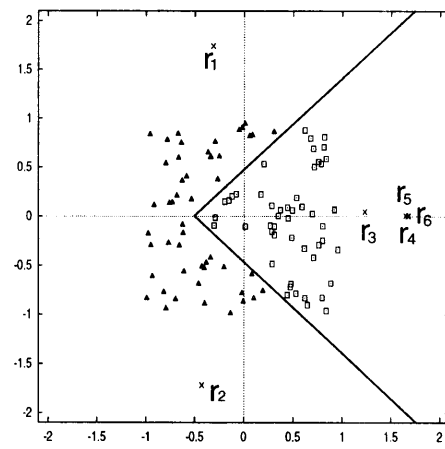
(a) 代表点 3 個



(b) 代表点 4 個



(a) 代表点 5 個



(b) 代表点 6 個

図 5.1: 代表点の配置 1

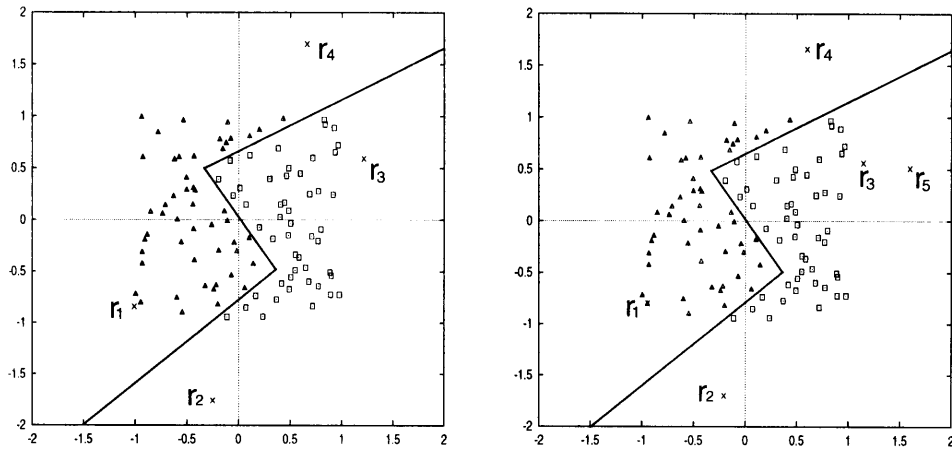
近に集まってから遠ざかる。従って結果は代表点の初期配置にはあまり関係しない。次に識別境界が図 5.2 のような 100 個のデータでやってみた。この例では代表点は左のカテゴリーに 2 個、右のカテゴリーに 2 個で必要十分である。代表点の総数を 4 個、5 個、6 個としたときの結果はそれぞれ図 5.2 の (a), (b), (c) のようになった。図 (b) では r_5 が、(c) では r_3 と r_6 が余り、これらの余剰代表点にはどのデータも所属しなかった。以上のように本学習法によれば十分多くの代表点を用意しておけば、各カテゴリーに必要最小限の数の代表点が割り振られ、残りの代表点はどのデータにも割り当てられずに廃棄されるという結果が得られた。また次節で扱うデータ例についても同様な代表点の割り当てが観測された [64]。

5.4.2 関数の次元

最近傍識別器の学習で用いる関数の次元やそれが各カテゴリーでとる値は任意に設定することができるが、その設定は学習の効率に影響するであろう。最もポピュラーなのは関数の次元をカテゴリーの数と同じにして関数の各成分は自分に対応するカテゴリーでだけ値 1 をとり、その他のカテゴリーでは 0 となるとするものである。これが最も次元が高い設定（これより高い次元は無駄である）であるが、最も低い次元としては関数を 1 次元として、各カテゴリーでカテゴリーの番号と同じ値をとるとするものがあり、これらの両極端の中間としていろいろな設定があり得る。そこで例題として簡単な時系列の識別についてこのような関数の設定の影響について調べてみた。 x を $d_i = -1, -0.8, -0.6, \dots, 0.8, 1$ の 11 点とり、波形 $y_i = f(d_i) + \epsilon_i$ を作った。 ϵ_i は $[-0.1, 0.1]$ の正規乱数とした。カテゴリーは四つとし、1 番目のカテゴリーでは $f(d) = -x$ 、2 番目では

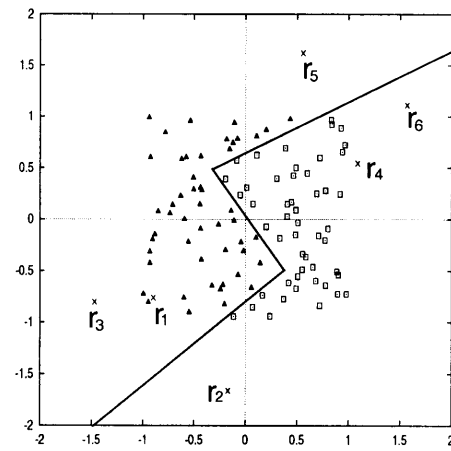
$$f(d) = \begin{cases} d + 1 & (d < -0.5) \\ -d & (-0.5 \leq d < 0.5) \\ d - 1 & (d \geq 0.5) \end{cases} \quad (5.21)$$

とし、3 番目では $f(d) = x$ 、4 番目では (5.21) の符号を反転したものとした。ノイズを変えて各カテゴリーに 5 個データを用意した。すなわちデータは 11 次元ベクトル (y_1, \dots, y_{11}) であり、各カテゴリーに 5 個ずつである。2 段木探索クラスタリング回



(a) 代表点 4 個

(b) 代表点 5 個



(c) 代表点 6 個

図 5.2: 代表点の配置 2

帰で学習を行った。

まず関数の次元を2次元とした場合の結果を表5.1に示す。表の0という番号をふつてある最初の行では第1,2,3,4カテゴリーにそれぞれ(0,0), (0,1), (1,0), (1,1)という関数値を割り当てた。右2行での○印は順当な階層分割すなわち1段目で第1と第2カテゴリーが一つにまとめられ、第3と第4が別のグループにまとめられ、2段目でこれらが各々2分割されたことを表す。△印はそれ以外の階層分割がなされたことを表す。ただしその場合でも2段目では正しく四つのカテゴリーに分割されている。このように関数が2次元の場合にはただ一つ(18番目)の例で一括学習が逐次学習よりも自然な階層分割になるという差がある程度でほとんど違いはない。次に関数を1次元にして各カテゴリーでとる値を0,1,2,3にしてみた結果を表5.2に示す。表中の×印は2段目で4個に分割できなかったことを表す。実際には×印の場合では全て1段目で三つのカテゴリーと一つのカテゴリーとに大分割されていた。このように関数が1次元の場合には逐次学習では学習に失敗するケースが多い。従って一括学習の方が性能がよいといえる。

5.5 むすび

クラスタリング関数回帰による最近傍識別器の教師あり学習法を提案して、十分な数の代表点を用意しておけば各カテゴリーに必要な十分な数の代表点が自動的に割り振られることや木探索識別器の代表点を一括して最適化する学習が行えることなどを簡単な例で確認した。本学習法は温度の自動アニーリングも内蔵するので誤差の少ない最適解を得ることができるが、収束に時間がかかるので高速化を図る必要がある。また本方法は十分多くの代表点を用意してそれが刈り込まれるタイプの学習法であるが、少数の代表点から始めて次第に数を増やしていくタイプの方が効率がよいと思われる。このような他のタイプの学習法の検討も今後の課題である。

表 5.1: 2次元関数による分割の結果

	1	2	3	4	一括	逐次
0	(0,0)	(1,0)	(0,1)	(1,1)	○	○
1	(0,0)	(1,0)	(1,1)	(0,1)	○	○
2	(0,0)	(0,1)	(1,0)	(1,1)	○	○
3	(0,0)	(0,1)	(1,1)	(1,0)	○	○
4	(0,0)	(1,1)	(1,0)	(0,1)	△	△
5	(0,0)	(1,1)	(0,1)	(1,0)	△	△
6	(1,0)	(0,0)	(0,1)	(1,1)	○	○
7	(1,0)	(0,0)	(1,1)	(0,1)	○	○
8	(1,0)	(0,1)	(0,0)	(1,1)	△	△
9	(1,0)	(0,1)	(1,1)	(0,0)	△	△
10	(1,0)	(1,1)	(0,0)	(0,1)	○	○
11	(1,0)	(1,1)	(0,1)	(0,0)	○	○
12	(0,1)	(0,0)	(1,0)	(1,1)	○	○
13	(0,1)	(0,0)	(1,1)	(1,0)	○	○
14	(0,1)	(1,0)	(0,0)	(1,1)	△	△
15	(0,1)	(1,0)	(1,1)	(0,0)	△	△
16	(0,1)	(1,1)	(0,0)	(1,0)	○	○
17	(0,1)	(1,1)	(1,0)	(0,0)	○	○
18	(1,1)	(0,0)	(1,0)	(0,1)	△	△
19	(1,1)	(0,0)	(0,1)	(1,0)	○	△
20	(1,1)	(1,0)	(0,0)	(0,1)	○	○
21	(1,1)	(1,0)	(0,1)	(0,0)	○	○
22	(1,1)	(0,1)	(0,0)	(1,0)	○	○
23	(1,1)	(0,1)	(1,0)	(0,0)	○	○

表 5.2: 1次元関数による分割の結果

	1	2	3	4	一括	逐次
0	0	1	2	3	○	○
1	0	1	3	2	○	○
2	0	2	1	3	△	×
3	0	2	3	1	△	△
4	0	3	1	2	○	×
5	0	3	2	1	○	△
6	1	0	2	3	○	○
7	1	0	3	2	○	○
8	1	2	0	3	○	×
9	1	2	3	0	○	△
10	1	3	0	2	△	×
11	1	3	2	0	○	△
12	2	0	1	3	○	△
13	2	0	3	1	△	×
14	2	1	0	3	△	△
15	2	1	3	0	△	×
16	2	3	0	1	○	○
17	2	3	1	0	○	○
18	3	0	1	2	△	△
19	3	0	2	1	○	×
20	3	1	0	2	○	△
21	3	1	2	0	△	×
22	3	2	0	1	○	○
23	3	2	1	0	○	○

第 6 章

アフィン部分空間への射影による最近傍 識別器の学習法

6.1 まえがき

前章で述べた最近傍識別においてデータベクトルの次元を圧縮することも行われ、線形射影による場合は代表超平面となる。VQPCA[36] や ASSOM[37, 38] はそのような識別器の教師なし学習法である。また部分空間識別法 [39] は教師あり学習の代表例である。このような線形射影による次元圧縮では主成分分析がよく用いられ、顔の認識などにおいてそのような手法の実用性 [40, 41] や心理実験との合致性も確認されており [42]、さらにニューラルネットによる非線形射影にも発展されている [43, 44]。一方主成分分析は、データの再生誤差は小さいが必ずしも識別に最適な射影ではなく [45, 46]、識別を目的とする判別分析による射影も提案され、主成分分析よりも高い認識率が達成されている [47, 48]。また部分空間法においてカテゴリー相互の関係も取り入れた相対 KL 変換法 [49] などが提案されている。しかしこれらの改良案はいずれも認識率そのものを直接目的関数とする最適化ではなく間接的な、すなわち識別そのものではなく判別的な最適化問題として定式化されており、真に識別に最適な射影となっているかどうか定かではない。学習部分空間法 [50] は反復法により認識率を高めていく方法であるが、最適化問題として定式化されていないので真の目的関数が

はっきりしない。そこで本章では問題をより直接的に最適化問題として定式化した学習法を提案する。

6.2 アフィン射影最近傍パターン識別器の学習

ここではカテゴリーの代表を点からアフィン部分空間へ拡張する。アフィン射影最近傍識別器は、入力データを各アフィン部分空間に射影し、その残差が最も小さいカテゴリーに分類する。アフィン空間は $\{r+x \mid Px = x\}$ と表せる。 P は射影行列である。 r をこのアフィン空間の原点と呼ぶことにする。部分空間として、(1) 射影行列がカテゴリー毎に異なり、原点は全てのカテゴリーで同じ、(2) 射影行列が全てのカテゴリーで同じであり、原点がカテゴリー毎に異なる、(3) 射影行列も原点もカテゴリー毎に異なる、の3通りを考える。

6.2.1 教師なし学習

まずアフィン射影識別器の教師なし学習の定式化を行なう。例えば射影行列も原点もカテゴリー毎に異なる場合には次式のように定式化される。

$$\begin{aligned} \min_{x_{ij} \in \{0,1\}, r_j, W_j} \quad & \sum_{i=1}^m \sum_{j=1}^n x_{ij} D_{ij} \\ \text{subj.to} \quad & \sum_{j=1}^n x_{ij} = 1 \end{aligned} \quad (6.1)$$

ここで D_{ij} はデータ d_i と第 j 部分空間との距離であり

$$D_{ij} = \left\| d_i - r_j - W_j W_j^T (d_i - r_j) \right\|^2 \quad (6.2)$$

である。式(6.1)の解は x_{ij} を

$$x_{ij} = \frac{e^{-D_{ij}/T}}{\sum_{k=1}^n e^{-D_{ik}/T}} \quad (6.3)$$

として r_j と W_j とを

$$\frac{dr_j}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial r_j} D_{ik} - \sum_{i=1}^m x_{ij} \frac{\partial D_{ij}}{\partial r_j} \quad (6.4.1)$$

$$\frac{\partial x_{ik}}{\partial r_j} = \begin{cases} \frac{1}{T} \frac{\partial D_{ij}}{\partial r_j} x_{ij} (x_{ij} - 1) & , k = j \\ \frac{1}{T} \frac{\partial D_{ij}}{\partial r_j} x_{ik} x_{ij} & , k \neq j \end{cases} \quad (6.4.2)$$

$$\frac{\partial D_{ij}}{\partial r_j} = -2(I - W_j^T W_j) e_{ij} \quad (6.4.3)$$

$$\frac{dW_j}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial W_j} D_{ik} - \sum_{i=1}^m x_{ij} \frac{\partial D_{ij}}{\partial W_j} \quad (6.4.4)$$

$$\frac{\partial x_{ik}}{\partial W_j} = \begin{cases} \frac{1}{T} \frac{\partial D_{ij}}{\partial W_j} x_{ij} (x_{ij} - 1) & , k = j \\ \frac{1}{T} \frac{\partial D_{ij}}{\partial W_j} x_{ik} x_{ij} & , k \neq j \end{cases} \quad (6.4.5)$$

$$\frac{\partial D_{ij}}{\partial W_j} = (d_i - r_j) e_{ij}^T W_j + e_{ij} (d_i - r_j)^T W_j \quad (6.4.6)$$

$$e_{ij} = (d_i - r_j) - W_j W_j^T (d_i - r_j) \quad (6.4.7)$$

の平衡解として求めながら T を 0 まで減少させるアニーリングによって得られる。
以下上記の 3 通りの場合それぞれについて実験を行う。

6.2.1.1 射影行列がカテゴリ毎に異なり、原点は全てのカテゴリで同じ場合

この場合第 j 部分空間は $\{p_j \mid p_j = r + q_j, W_j W_j^T q_j = q_j\}$ であり、以下の最適化問題として定式化される。

$$\begin{aligned} \min_{x_{ij} \in \{0,1\}, r, W_j} & \sum_{i=1}^m \sum_{j=1}^n x_{ij} D_{ij} \\ \text{subj.to} & \sum_{j=1}^n x_{ij} = 1 \end{aligned} \quad (6.5)$$

ここで D_{ij} は

$$D_{ij} = \left\| d_i - r - W_j W_j^T (d_i - r) \right\|^2 \quad (6.6)$$

となる。式 (6.5) の解は x_{ij} を 6.3 として r と W_j とを

$$\frac{dr}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial r} D_{ik} - \sum_{i=1}^m x_{ij} \frac{\partial D_{ij}}{\partial r} \quad (6.7.1)$$

$$\frac{\partial x_{ik}}{\partial r} = - \left(\frac{1}{T} \frac{\partial D_{ik}}{\partial r} - \frac{\sum_{l=1}^n \frac{1}{T} \frac{\partial D_{il}}{\partial r} e^{-D_{il}/T}}{\sum_{l=1}^n e^{-D_{il}/T}} \right) x_{ik} \quad (6.7.2)$$

$$\frac{\partial D_{ij}}{\partial r} = -2(I - W_j^T W_j) e_{ij} \quad (6.7.3)$$

$$\frac{dW_j}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial W_j} D_{ik} - \sum_{i=1}^m x_{ij} \frac{\partial D_{ij}}{\partial W_j} \quad (6.7.4)$$

$$\frac{\partial x_{ik}}{\partial W_j} = \begin{cases} \frac{1}{T} \frac{\partial D_{ij}}{\partial W_j} x_{ij} (x_{ij} - 1) & , k = j \\ \frac{1}{T} \frac{\partial D_{ij}}{\partial W_j} x_{ik} x_{ij} & , k \neq j \end{cases} \quad (6.7.5)$$

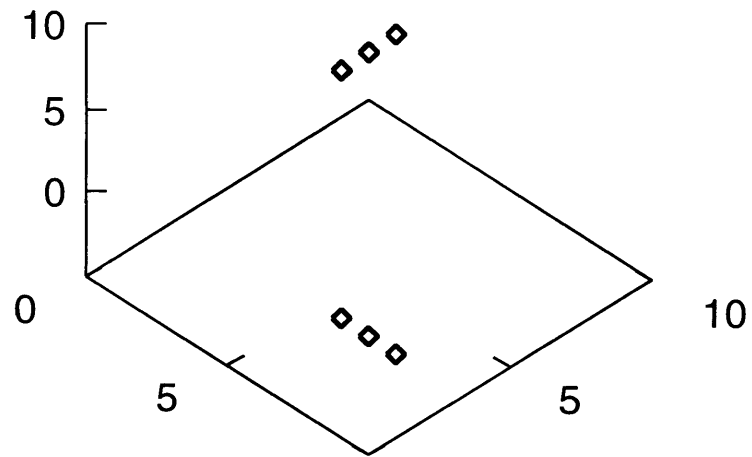
$$\frac{\partial D_{ij}}{\partial W_j} = (d_i - r) e_{ij}^T W_j + e_{ij} (d_i - r)^T W_j \quad (6.7.6)$$

$$e_{ij} = (d_i - r) - W_j m W_j^T (d_i - r) \quad (6.7.7)$$

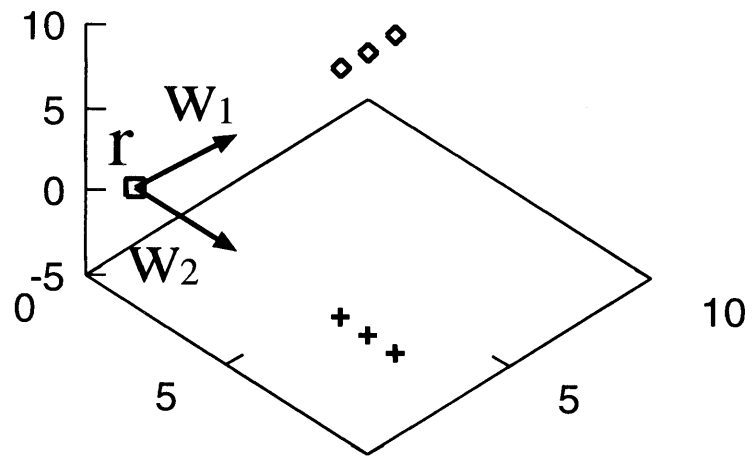
の平衡解として求めながら T を 0 まで減少させるアニーリングによって得られる。

この場合の典型的なデータで学習を行った結果を図 6.1 に示す。図 6.1(1) に示すように、それぞれ違う方向の直線上に並んだ 3 次元の点列を学習データとし、1 次元部分空間に射影し 2 カテゴリーにクラスタリングした。図 6.1(2) は学習後の結果である。 r は原点、矢印は部分空間の 1 次元基底ベクトル、点列はそれらが属するカテゴリーの部分空間に射影したものである。学習された基底が点列の方向を向いている。このようにこの場合は、原点から見てどの方向にデータが存在するかによって分類が行われる。

次に吉川[51] による図形の分類の心理実験に使用された顔のデータセットを用いて学習を行った結果を図 6.2 に示す。このデータは、顔図形において両目間の距離、目の高さ、口の高さの 3 つの属性が 9 段階に変化する 26 個のデータよりなる。この 3 次元データを 1 次元部分空間に射影し 2 カテゴリーにクラスタリングした。その結果 15.2% の誤分類が発生した。

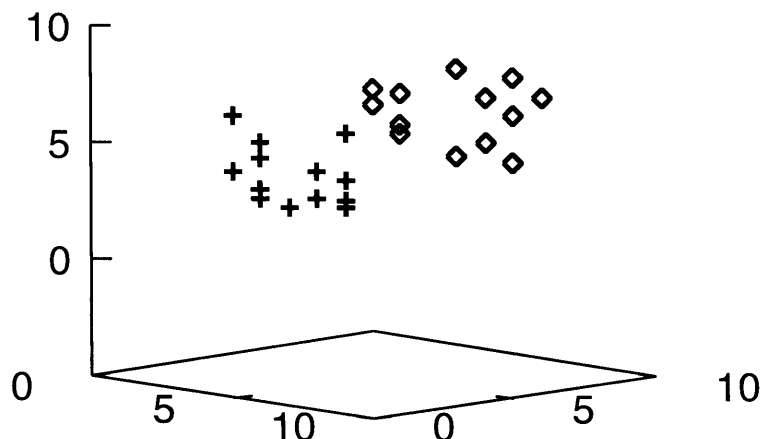


(1)input data

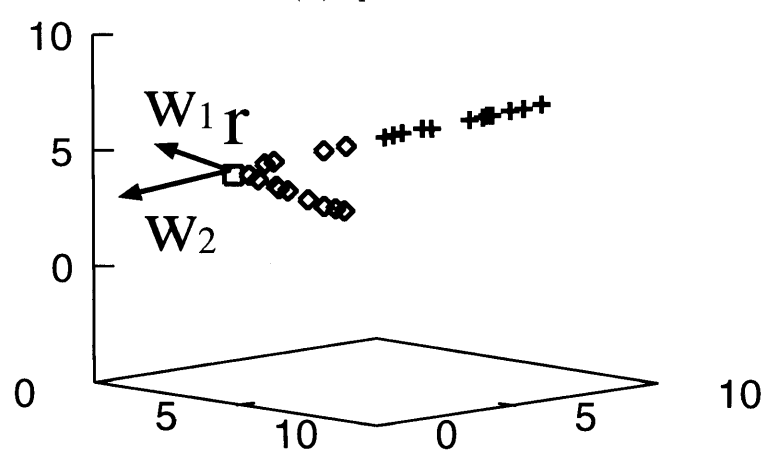


(2)subspaces

図 6.1: 簡単なデータによる学習結果 (教師なし 1)



(1)input data



(2)subspaces

図 6.2: 顔データによる学習結果 (教師なし1)

6.2.1.2 射影行列が全てのカテゴリで同じであり，原点がカテゴリ毎に異なる場合

この場合第 j 部分空間は $\{p_j \mid p_j = r_j + q, WW^T q = q\}$ であり，以下の最適化問題として定式化される．

$$\begin{aligned} \min_{x_{ij} \in \{0,1\}, r_j, W} & \sum_{i=1}^m \sum_{j=1}^n x_{ij} D_{ij} \\ \text{subj.to} & \sum_{j=1}^n x_{ij} = 1 \end{aligned} \quad (6.8)$$

ここで D_{ij} は

$$D_{ij} = \|d_i - r_j - WW^T(d_i - r_j)\|^2 \quad (6.9)$$

となる．式(6.5)の解は x_{ij} を6.3として r_j と W とを

$$\frac{dr_j}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial r_j} D_{ik} - \sum_{i=1}^m x_{ij} \frac{\partial D_{ij}}{\partial r_j} \quad (6.10.1)$$

$$\frac{\partial x_{ik}}{\partial r_j} = \begin{cases} \frac{1}{T} \frac{\partial D_{ij}}{\partial r_j} x_{ij} (x_{ij} - 1) & , k = j \\ \frac{1}{T} \frac{\partial D_{ij}}{\partial r_j} x_{ik} x_{ij} & , k \neq j \end{cases} \quad (6.10.2)$$

$$\frac{\partial D_{ij}}{\partial r_j} = -2(I - W^T W) e_{ij} \quad (6.10.3)$$

$$\frac{dW}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial W} D_{ik} - \sum_{i=1}^m x_{ij} \frac{\partial D_{ij}}{\partial W} \quad (6.10.4)$$

$$\frac{\partial x_{ik}}{\partial W} = - \left(\frac{1}{T} \frac{\partial D_{ik}}{\partial W} - \frac{\sum_{l=1}^n \frac{1}{T} \frac{\partial D_{il}}{\partial W} e^{-D_{il}/T}}{\sum_{l=1}^n e^{-D_{il}/T}} \right) x_{ik} \quad (6.10.5)$$

$$\frac{\partial D_{ij}}{\partial W} = (d_i - r_j) e_{ij}^T W + e_{ij} (d_i - r_j)^T W \quad (6.10.6)$$

$$e_{ij} = (d_i - r_j) - WW^T(d_i - r_j) \quad (6.10.7)$$

の平衡解として求めながら T を0まで減少させるアニーリングによって得られる．

典型的なデータに対して学習を行った結果を図6.3に示す．図6.3(1)に示すように，平行に並んだ点列を学習データとし，1次元部分空間に射影し2カテゴリにクラスタリングした．図6.3(2)は学習後の結果である．この2つの点は学習データを部分空間に射影したものおよびカテゴリの代表点 r_j であり重なっている．

次に上記の吉川による顔のデータセットを用いて学習を行った結果を図 6.4 に示す。この結果では学習データに対して 100% の識別率が得られた。

6.2.1.3 射影行列も原点もカテゴリー毎に異なる場合

この場合第 j 部分空間は $\{p_j \mid p_j = r_j + q_j, W_j W_j^T q_j = q_j\}$ であり D_{ij} は式 (6.2) であり, r_j と W_j は式 (6.10) で求める。典型的なデータに対して学習を行った結果を図 6.5 に示す。2 つの 2 次元平面上に並んだ 3 次元の点列と 1 つの直線上に並んだ 3 次元の点列を学習データとし, 2 次元部分空間に射影するカテゴリーを 2 つ, 1 次元部分空間に射影するカテゴリーを 1 つとしてクラスタリングした。学習後の原点 r と部分空間の基底を矢印で示している。このように学習データの各カテゴリー毎の次元と同じ次元の部分空間がそれぞれ自動的に割り当てられた。

次に上記の吉川による顔のデータセットを用いて学習を行った結果を図 6.6 に示す。この場合では 3.8% の誤分類が発生した。

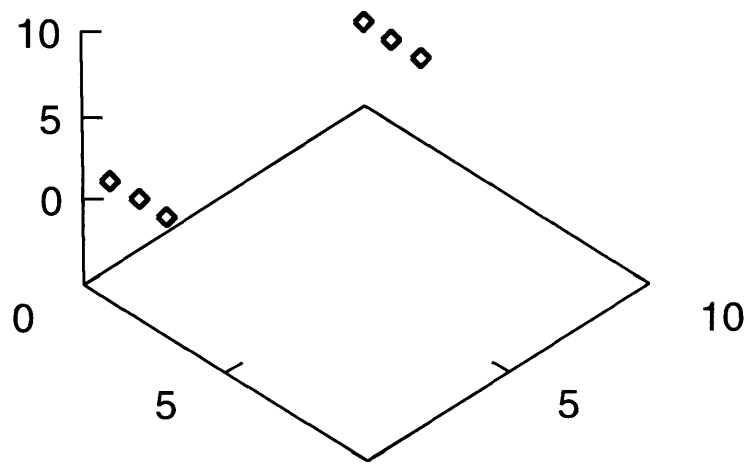
6.2.2 教師あり学習

式 (6.1) において x_{ij} を,

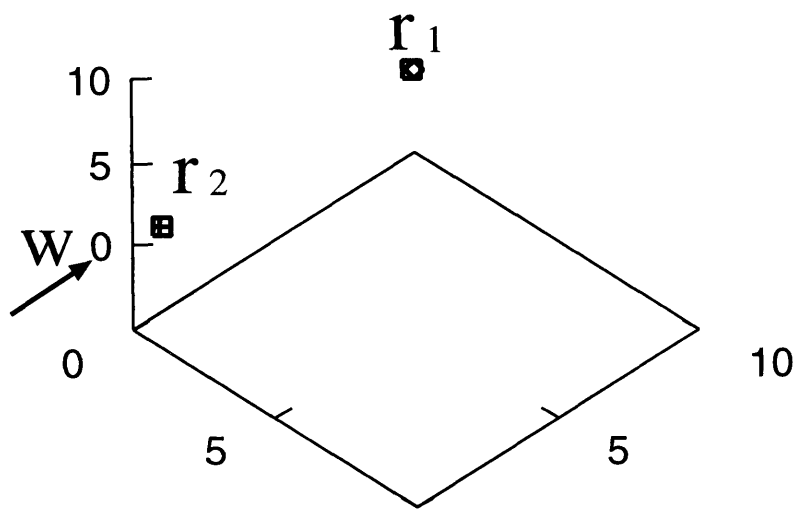
$$x_{ij} = \begin{cases} 1 & d_i \in \text{第 } j \text{ カテゴリー} \\ 0 & \text{その他} \end{cases} \quad (6.11)$$

と固定すると, カテゴリー毎に互いに独立に部分空間を求めることになる。これまでこのような学習法が多く用いられている [52]。

簡単なデータについてこの学習法で部分空間を求めた結果を図 6.7 に示す。左上の図が使用したデータである。8 個のデータのうち上 3 個が 1 つのグループでこれらは直線に並んでおり, 残る下 5 個がもう 1 つのグループで, 矢印で示す左端の点は上 3 個の直線上にあり下 4 個の直線からは少しずれている。これらを 1 次元部分空間に射影し 2 カテゴリーに分類するように学習させた結果が図 6.7 の残り 3 つである。これらはそれぞれ, (1) 射影行列がカテゴリー毎に異なり, 原点は全てのカテゴリーで同じ, (2) 射影行列が全てのカテゴリーで同じであり, 原点がカテゴリー毎に異なる,



(1)input data



(2)subspaces

図 6.3: 簡単なデータによる学習結果 (教師なし 2)

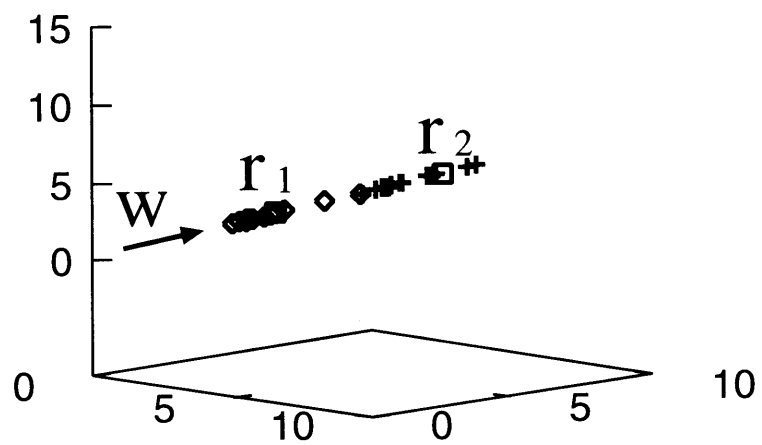


図 6.4: 顔データによる学習結果 (教師なし 2)

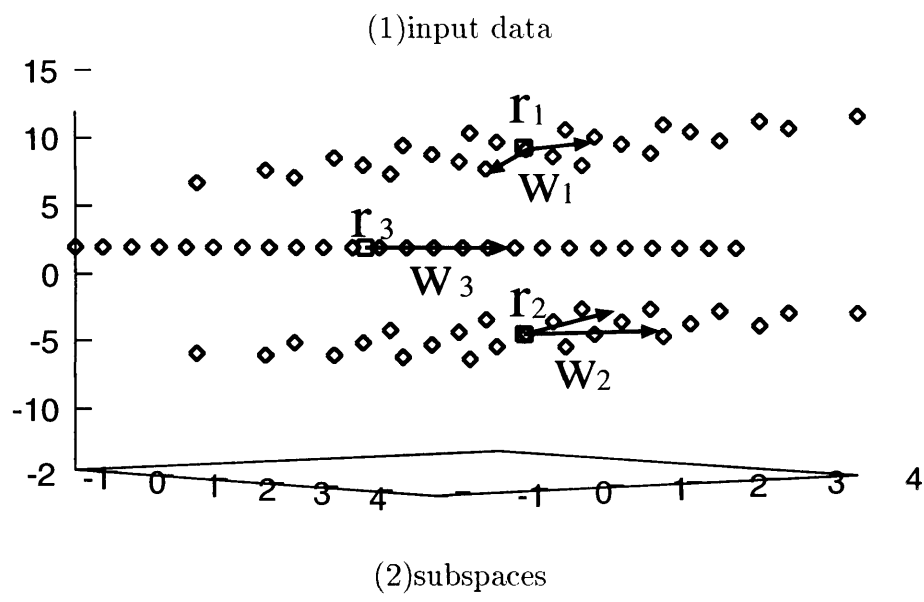
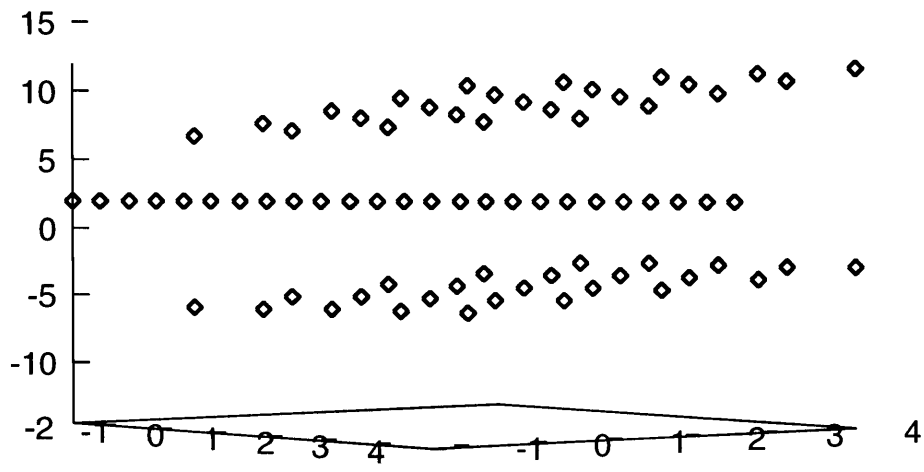


図 6.5: 簡単なデータによる学習結果 (教師なし 3)

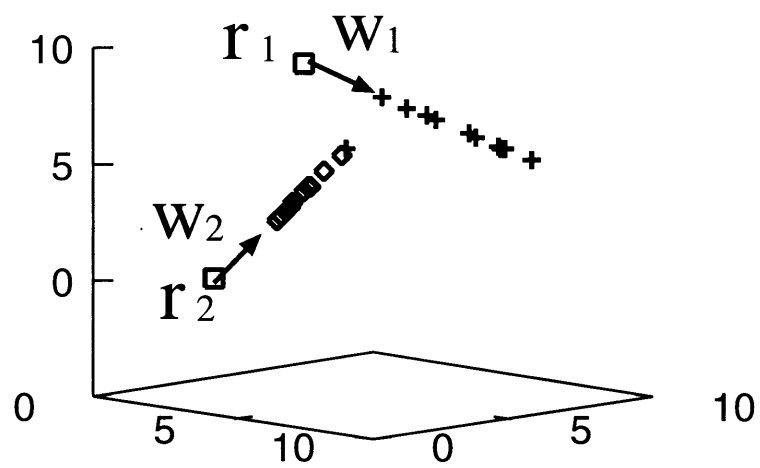


図 6.6: 顔データによる学習結果 (教師なし 3)

(3) 射影行列も原点もカテゴリー毎に異なる，場合である．この結果，(1),(2),(3)の場合全てで左端の点が第1カテゴリーに誤分類され12.5%の誤分類が発生した．次に上記の吉川による顔のデータセットを用いて同じ方法で学習を行った結果を図6.8に示す．このデータでは3つの場合全てで3.8%の誤分類が発生した．

このようにカテゴリー毎に独立に部分空間を求めると，個々のカテゴリー内の特徴は抽出されるがカテゴリー間の関係は考慮されないため，識別にとって最適な学習ではない．カテゴリー間の関係を考慮する学習法として相対KL変換法[49]が提案されているが，カテゴリー間の抑制をどの程度かけるかが明示的でない．そこで，ここでは教師ありアフィン射影識別を式(5.10)と同じく以下のように定式化する．

$$\min_{r_j, W_j, g_j} \sum_{i=1}^m \sum_{j=1}^n x_{ij} (f_i - g_j)^2 \quad (6.12)$$

ここで x_{ij} は，式(6.3)を用いる．この式の解は

$$\frac{dr_j}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial r_j} (f_i - g_k)^2 \quad (6.13.1)$$

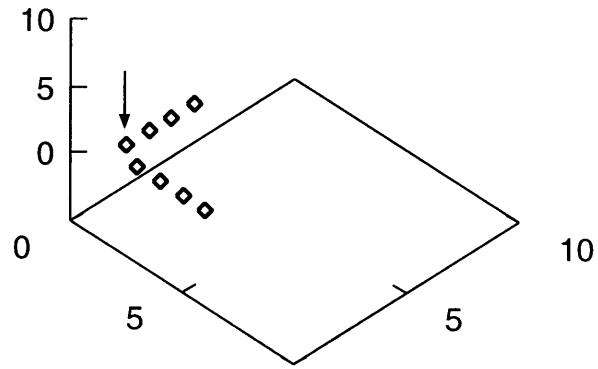
$$\frac{dW_j}{dt} = - \sum_{i=1}^m \sum_{k=1}^n \frac{\partial x_{ik}}{\partial W_j} (f_i - g_k)^2 \quad (6.13.2)$$

$$\frac{dg_j}{dt} = 2 \sum_{i=1}^m x_{ij} (f_i - g_j) \quad (6.13.3)$$

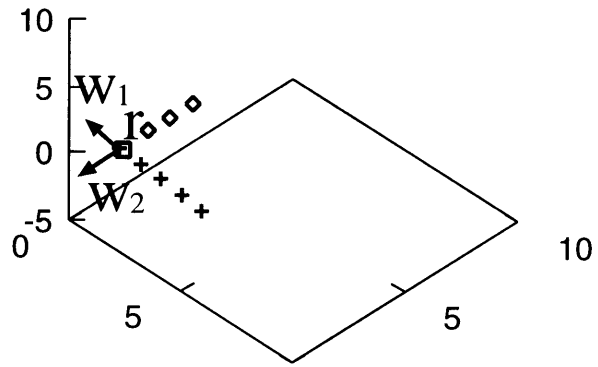
の平衡解を求めることにより得られる．ここで $\partial x_{ik}/\partial r_j$ と $\partial x_{ik}/\partial W_j$ は式(6.10)で与えられている．教師なしの場合と同様に T はアニーリングする．教師ありの場合も，教師なしの時と同様に部分空間の構成として前述の3つの場合が考えられる．以下これら3通りの場合それぞれについて実験を行う．

6.2.2.1 射影行列がカテゴリー毎に異なり，原点は全てのカテゴリーで同じ場合

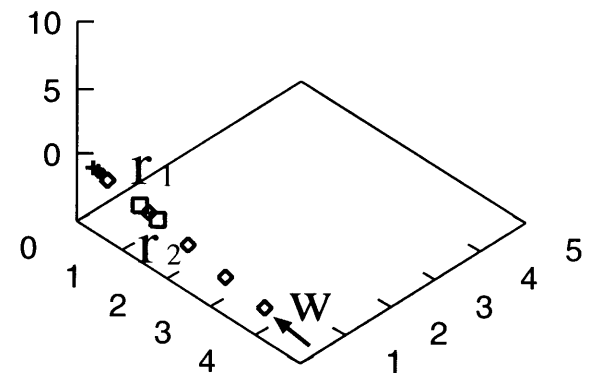
この場合 D_{ij} は式(6.6)であり， r, W_j, g_j は(6.13.1)で r_j を r としたものと(6.13.2), (6.13.3)から得られる．図6.7のデータに対してこの方法で学習を行った結果を図6.9に示す．入力データを1次元部分空間に射影し2カテゴリーに分類した．この結果，学習データに対して100%の識別率が得られ，カテゴリー毎に独立に部分空間を求める教師あり学習(図6.7(1))よりも改善された．



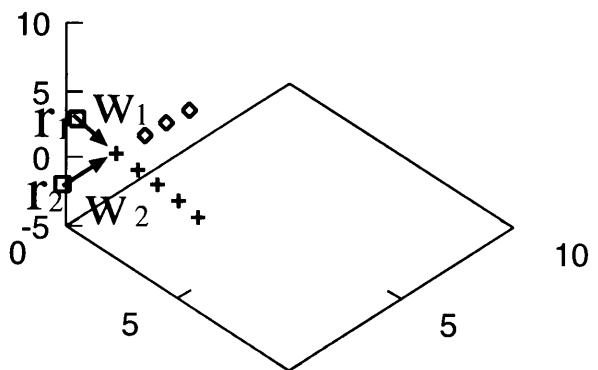
input data



(1)subspaces

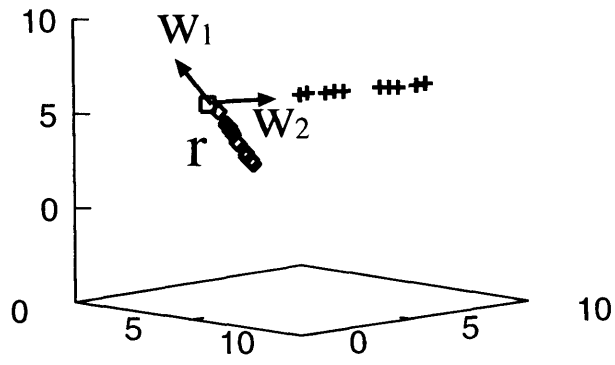


(2)subspaces

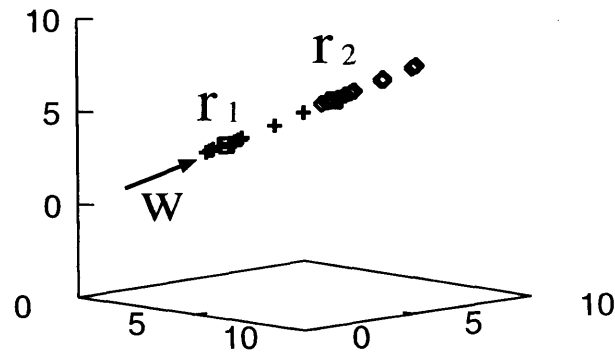


(3)subspaces

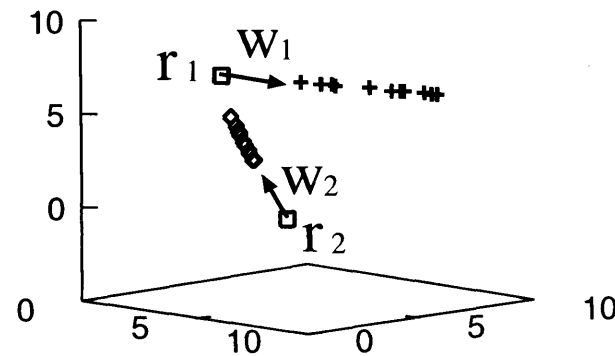
図 6.7: 簡単なデータによる学習結果 (CLAFIC)



(1)subspaces



(2)subspaces



(3)subspaces

図 6.8: 顔データによる学習結果 (CLAFIC)

次に上記の吉川による顔のデータセットを用いて学習を行った結果を図 6.10 に示す。この結果、学習データに対して 100% の識別率が得られ、3.1.1 節の教師なしの場合やカテゴリー毎に部分空間を求める教師あり学習 (図 6.8(1)) よりも改善された。

6.2.2.2 射影行列が全てのカテゴリーで同じであり、原点がカテゴリー毎に異なる場合

この場合 D_{ij} は式 (6.9) であり、 r_j, W, g_j は (6.13.2) で W_j を W としたものと (6.13.1), (6.13.3) から得られる。図 6.7 のデータに対する結果を図 6.11 に示す。入力データを 1 次元部分空間に射影し 2 カテゴリーに分類した。この結果学習データに対して 100% の識別率が得られ、カテゴリー毎に部分空間を求める教師あり学習 (図 6.7(2)) よりも改善された。

次に上記の吉川による顔のデータセットを用いて学習を行った結果を図 6.12 に示す。この結果学習データに対して 100% の識別率が得られ、カテゴリー毎に部分空間を求める教師あり学習 (図 6.8(2)) よりも改善された。

6.2.2.3 射影行列も原点もカテゴリー毎に異なる場合

この場合 D_{ij} は式 (6.2) であり、 r_j と W_j は式 (6.17) で求める。図 6.7 のデータに対する結果を図 6.13 に示す。入力データを 1 次元部分空間に射影し 2 カテゴリーに分類した。この結果学習データに対して 100% の識別率が得られ、カテゴリー毎に部分空間を求める教師あり学習 (図 6.7(3)) よりも改善された。

次に上記の吉川による顔のデータセットを用いて学習を行った結果を図 6.14 に示す。この結果学習データに対して 100% の識別率が得られ、3.1.3 節の教師なしの場合やカテゴリー毎に部分空間を求める教師あり学習 (図 6.8(3)) よりも改善された。

6.3 むすび

本章では、アフィン部分空間最近傍識別器の教師なし、及び教師ありの勾配法による学習法を提案し、簡単な例題による実験でその性能を確認した。教師なし学習では

データの分布に即したクラスタリングが得られ、教師あり学習ではカテゴリー毎に独立に部分空間を求める方法よりも高い識別率が得られた。本方法はアニーリングを用いているため誤差の小さい解が得られる反面、学習に時間がかかる。今後学習速度の向上を図り、実画像データ分類などの高次元で大規模なデータを用いた実験を行いたい。

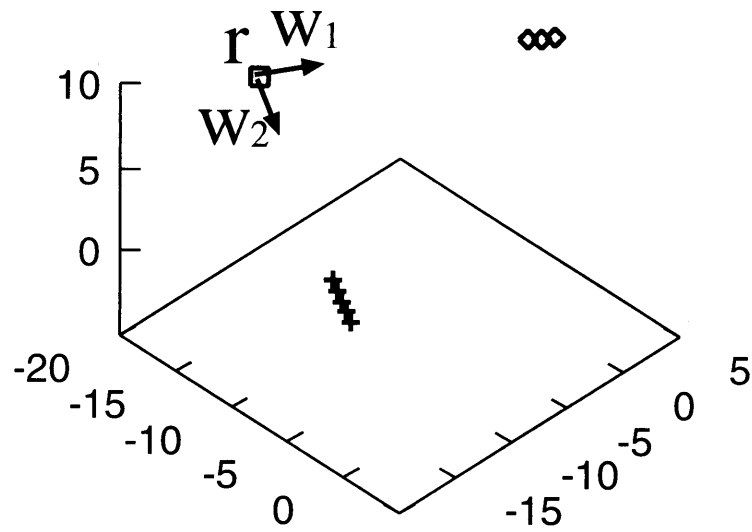


図 6.9: 簡単なデータによる学習結果 (教師あり 1)

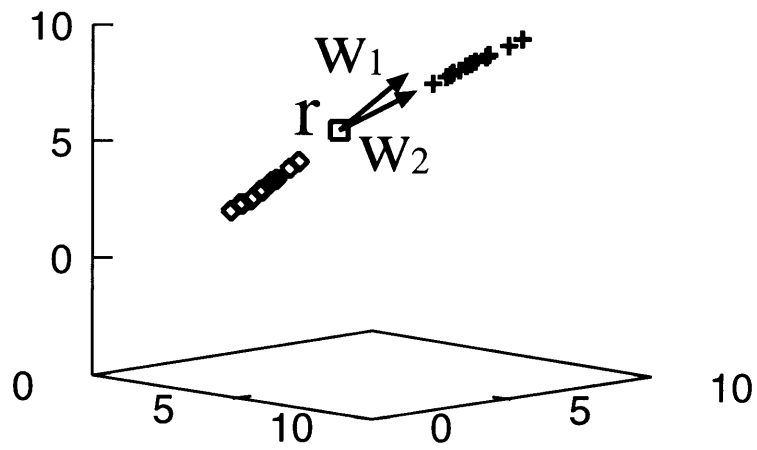


図 6.10: 顔データによる学習結果 (教師あり 1)

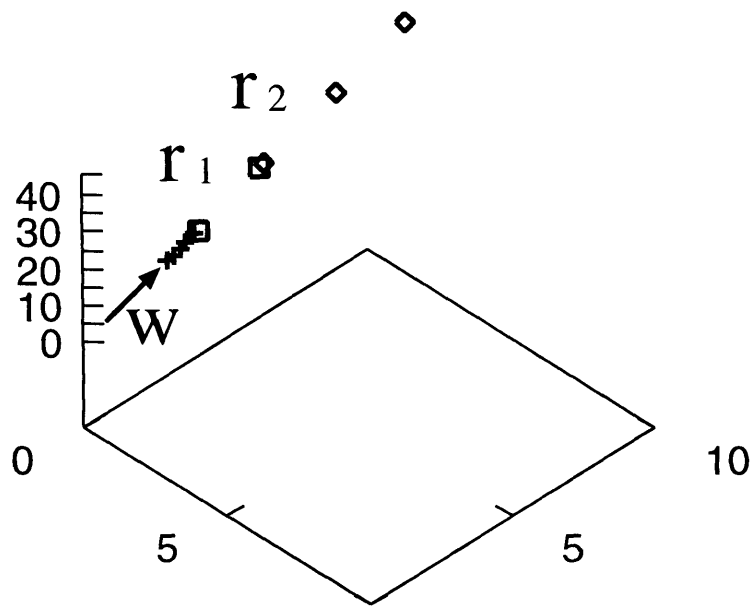


図 6.11: 簡単なデータによる学習結果 (教師あり 2)

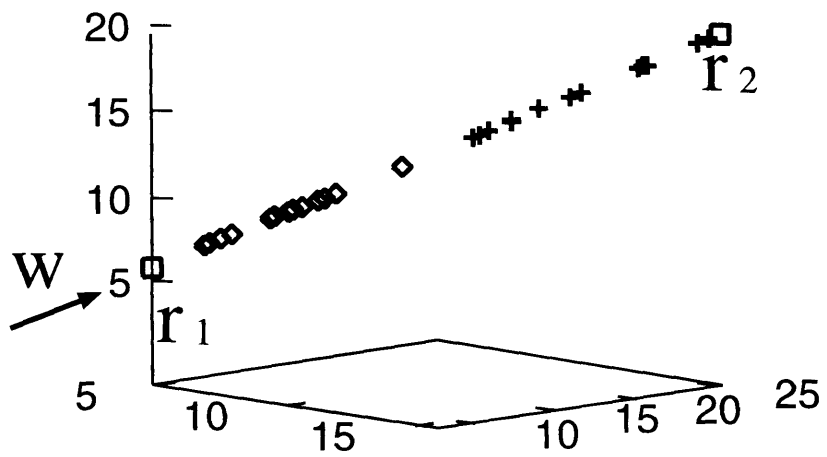


図 6.12: 顔データによる学習結果 (教師あり 2)

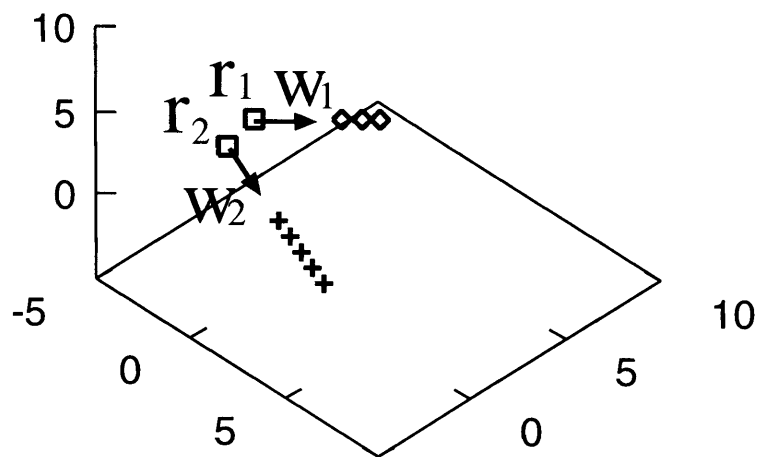


図 6.13: 簡単なデータによる学習結果 (教師あり 3)

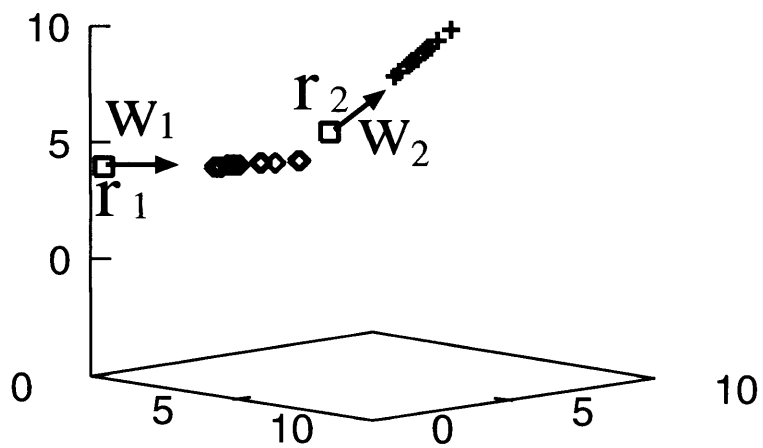


図 6.14: 顔データによる学習結果 (教師あり 3)

第 7 章

最近傍識別器の教師あり / なし混合学習

7.1 まえがき

これまでに述べたように最近傍識別器の学習は、代表点の最適な配置を学習データに基づいて求める過程であり、教師なし学習と教師あり学習とに大別される。教師あり学習は識別誤差を最小化するように代表点を配置するものであり、カテゴリーラベル付きの学習データを用いて行われる。一方教師なし学習は量子化誤差を最小にするように代表点を配置するものであり、ラベルなしの（ラベルが付いていても無視して）学習データを用いて行われる。学習後の識別器は、教師あり学習では専ら識別に使われ、教師なし学習では識別に使われることもあるし、ベクトル量子化に使われることもある。これらの学習はどちらも代表点のある種の最適な配置を求めるものであるが、最適化の目的関数が異なるので得られる配置は一般的に異なり、一方の学習結果は他方の目的にとっては最適とはならない。本章ではこれらの学習を混在させた学習について考える。混在させる目的としては、識別器を同時にベクトル量子化器としても使う場合 [53] において、量子化誤差と識別誤差の両方を小さくしたい場合や、教師データのカテゴリーラベルの収集が困難で、少数のデータについてしかカテゴリーの情報が得られないときに、ラベルが未知のデータも加えて混合学習して般化能力の向上を図る場合 [54, 55, 56] などがあげられる。本章では勾配法を使う学習法についてこのような混合学習の方法を提案し、本学習法の性質を解析して、簡単なデータ例と顔の識

別の例で本方法の有効性を検証する。

7.2 教師あり / なし混合学習

以上の2種類の学習はどちらも代表点 r_j の最適な配置を求めるものであるが、最適化の目的関数が異なるので得られる配置は一般的に異なり、一方の学習結果は他方の目的にとっては最適とはならない。ここではこれらの学習を混在させた学習について考える。与えられるデータには全てラベルが付いているものとする。しかし学習に際してはこのうちのいくつか(0個も含む)のデータのラベルは未知とされるとする。学習にラベルが使われるデータを教師ありデータ、それ以外を教師なしデータということにする。このような混合データに対して教師あり / なし混合学習を行うのであるが、学習後の識別器の性能評価として、1) 全データの量子化誤差と教師ありデータの識別率の両方を小さくする、というのと、2) 全データの識別率を小さくする、という2通りを考える。前者は識別器を同時にベクトル量子化器としても使う場合 [53] に適用される。これは多目的最適化問題 [57] であり、式 (5.3) と (5.10) を組み合わせた

$$\min_{r_j, g_j} \left[\begin{array}{l} \sum_{i=1}^m \sum_{j=1}^n x_j(d_i) \|d_i - r_j\|^2 \\ \sum_{i \in I_s} \sum_{j=1}^n x_j(d_i) \|f_i - g_j\|^2 \end{array} \right] \quad (7.1)$$

となる。 I_s は教師ありデータの集合である。一方後者は、教師ありデータがごく少数のとき、教師なし学習も組み入れた方が般化能力が向上するという事 [54, 55] に関連する。これは、教師あり学習だけでは解が1意に定まらず、教師なし学習によりこの中から1つの解を選ぶことにより、全データの識別率の期待値が向上するものと予想される。これは制約付き最適化問題

$$\begin{array}{ll} \min_{r_j, g_j} & \sum_{i=1}^m \sum_{j=1}^n x_j(d_i) \|d_i - r_j\|^2 \\ \text{subj.to} & \sum_{i \in I_s} \sum_{j=1}^n x_j(d_i) \|f_i - g_j\|^2 = \epsilon \end{array} \quad (7.2)$$

で定式化される。 ϵ は教師あり学習の最適解での誤差である。以上の式 (7.1),(7.2) 両方とも混合学習

$$\min_{r_j, g_j} \sum_{i=1}^m \sum_{j=1}^n x_j(d_i) \|d_i - r_j\|^2 + \alpha \sum_{i \in I_s} \sum_{j=1}^n x_j(d_i) \|f_i - g_j\|^2 \quad (7.3)$$

に帰着できる。式 (7.3) の目的関数は、式 (7.1) の場合 2 つの目的関数を加えてスカラ化したものであり、式 (7.2) の場合ラグランジュ関数である。式 (7.3) で $\alpha \downarrow 0$ とするとラベルは無視して全てのデータについて教師なし学習をすることになり、逆に $\alpha \uparrow \infty$ とすると教師ありデータだけを用いて教師あり学習をすることになり、中間の α ではこれらの混合学習が行われる。このとき上記の性能評価の 1) に関して次の性質が成り立つ。

[性質 1] α を増やすと量子化誤差 $\sum \sum x_j(d_i) \|d_i - r_j\|^2$ は単調に増加し、教師ありデータの識別誤差 $\sum \sum x_j(d_i) \|f_i - g_j\|^2$ は単調に減少する。

(証明) $e_1 = \sum \sum x_j(d_i) \|d_i - r_j\|^2$, $e_2 = \sum \sum x_j(d_i) \|f_i - g_j\|^2$ とおくと式 (7.3) の解では目的関数 $e = e_1 + \alpha e_2$ の変数 $p = [r, g]$ に関する勾配は 0 すなわち

$$\frac{\partial e_1}{\partial p} + \alpha \frac{\partial e_2}{\partial p} = 0 \quad (7.4)$$

であり、ヘッセ行列 $\partial^2 e / \partial p^2 = \partial^2 e_1 / \partial p^2 + \alpha \partial^2 e_2 / \partial p^2$ は非負定値である。式 (7.4) を α で微分すると

$$\frac{\partial^2 e}{\partial p^2} \frac{dp}{d\alpha} + \frac{\partial e_2}{\partial p} = 0 \quad (7.5)$$

となり、これから

$$\frac{\partial e_2}{\partial p} = - \left(\frac{\partial^2 e}{\partial p^2} \right)^{-1} \frac{dp}{d\alpha} \quad (7.6)$$

を得る。これを

$$\frac{de_2}{d\alpha} = \frac{\partial e_2}{\partial p} \frac{dp}{d\alpha} \quad (7.7)$$

に代入すると

$$\frac{de_2}{d\alpha} = - \frac{dp}{d\alpha} \left(\frac{\partial^2 e}{\partial p^2} \right)^{-1} \frac{dp}{d\alpha} \quad (7.8)$$

となり、

$$\frac{\partial^2 e}{\partial p^2} \quad (7.9)$$

は非負定値であるから

$$\frac{de_2}{d\alpha} \leq 0 \quad (7.10)$$

となる。従って α を増やすと e_2 は単調に減少する。また式 (7.4) と (7.6) とから

$$\frac{\partial e_1}{\partial p} = \alpha \left(\frac{\partial^2 e}{\partial p^2} \right)^{-1} \frac{dp}{d\alpha} \quad (7.11)$$

となり、これを

$$\frac{de_1}{d\alpha} = \frac{\partial e_1}{\partial p} \frac{dp}{d\alpha} \quad (7.12)$$

に代入すると

$$\frac{de_1}{d\alpha} = \alpha \frac{dp}{d\alpha} \left(\frac{\partial^2 e}{\partial p^2} \right)^{-1} \frac{dp}{d\alpha} \quad (7.13)$$

となり、 $\alpha > 0$ であるので

$$\frac{de_1}{d\alpha} \geq 0 \quad (7.14)$$

となる。従って α を増やすと e_1 は単調に増加する。

(証明終)

このように性能評価の 1 に関しては、 α によって両誤差のトレードオフが調節されることになる。これに対し、性能評価の 2 については理論的な評価は困難である。定性的には、混合学習によって代表点のばらつきが小さくなり、それによって全データの識別率の期待値が増えることが期待される。これについて簡単に解析してみる。全データの識別誤差 E は代表点 $r = [r_1, \dots, r_n]$ の関数 $E(r)$ である。 $E(r)$ が最小になる r を r^* とする。 $E(r)$ を r^* の周りでテーラー展開すると

$$E(r) \doteq E(r^*) + g^T(r^*)(r - r^*) + \frac{1}{2} \text{tr}[H(r^*)(r - r^*)(r - r^*)^T] \quad (7.15)$$

となる。 g は勾配ベクトル、 H はヘッセ行列、 tr はトレースである。 r^* は最小点であるから $g(r^*) = 0, H(r^*) \geq 0$ である。従って $E(r)$ の期待値 $\bar{E}(r)$ は

$$\bar{E}(r) \doteq q + \frac{1}{2} \text{tr}[H(r^*)C(r)] \quad (7.16)$$

となる。 $C(r)$ は r の共分散行列で q は定数である。今 2 つの学習法があってそれぞれで得られる代表点を s と t とし、 $C(s) \geq C(t)$ すなわち $C(s) - C(t)$ が非負定値であるとする。するとヘッセ行列も非負定値であり、2 つの非負定値行列の積のトレース

ス是非負であるから $\bar{E}(s) \geq \bar{E}(t)$ となる。従って代表点のばらつきが小さいほど識別誤差も小さくなると期待される。そこで式(7.3)の混合学習について代表点の分散を簡単に評価してみる。式(7.3)で求まる代表点は式(7.4)の解すなわち

$$f_1(r) + \alpha f_2(d_s, r) = 0 \quad (7.17)$$

の解である。ここで $f_1(r) = \partial e_1 / \partial p$, $f_2(d_s, r) = \partial e_2 / \partial p$ であり、それぞれ代表点 r と教師ありデータ d_s の関数である。教師なしデータは試行によらず一定すなわち分散は0であるので考慮しなくてよい。式(7.15)の識別誤差 $E(r)$ を最小にする r^* を与える d_s を d_s^* とし、式(7.17)の f_1 と f_2 をこれらの周りでテーラー展開すると

$$F_{1r}(r^*)(r - r^*) + \alpha F_{2d_s}(d_s^*, r^*)(d_s - d_s^*) + \alpha F_{2r}(d_s^*, r^*)(r - r^*) \doteq 0 \quad (7.18)$$

を得る。ここで F_{1r} 等は $F_{1r} = \partial f_1 / \partial r$, $F_{2d_s} = \partial f_2 / \partial d_s$, $F_{2r} = \partial f_2 / \partial r$ というヤコビアン行列である。式(7.18)から

$$r - r^* = -\alpha(F_{1r} + \alpha F_{2r})^{-1} F_{2d_s}(d_s - d_s^*) \quad (7.19)$$

となる。ただし F_{1r} 等の引数は略した。式(7.19)から代表点の共分散行列 $C(r)$ は

$$C(r) = \alpha^2(F_{1r} + \alpha F_{2r})^{-1} \cdot F_{2d_s} C(d_s) F_{2d_s}^T (F_{1r} + \alpha F_{2r})^{-T} \quad (7.20)$$

となる。ここで $C(d_s)$ は d_s の共分散行列であり、 $-T$ は逆行列の転置を表す。式(7.20)から、 α すなわち教師あり学習の重みが小さいほど代表点のばらつきが小さくなると期待される。以上の解析結果から、教師ありデータが少数のときには、その少数の教師ありデータだけで教師あり学習をするよりも教師なしデータを加えて混合学習した方が、代表点の分散が小さくなり、従って識別誤差の期待値が小さくなると期待される。

7.3 実験例

簡単なデータの例を用いて以上の理論的な予想を実験で検証してみた。まず最初のデータは Duda と Hart の本 [60] に出てくる図 7.1 のようなものである。データは全部

で 43 個で右の 3 個が 1 つのカテゴリーで、残りの左の 40 個の固まりがもう 1 つのカテゴリーである。代表点は 2 個とした。実験では全て各カテゴリーに 1 個ずつ代表点が割り当てられた。教師あり学習での識別関数は 1 次元とし教師値は 0 と 1 とした。まず性質 1 を検証した。全データを教師ありデータとして混合学習した結果、性質 1 の通り e_1 と e_2 は α によって図 7.2 のように単調に変化した。次に混合学習による識別率の期待値の向上について実験してみた。まず図 7.3 は全データで教師なし学習をした結果である。このように識別境界がカテゴリー 1 に入り込み、識別率は 62.8% となった。次に図 7.4 は全データで教師あり学習をした結果であり、100% の識別率となった。次に教師ありデータの数が極端に少ない教師あり学習として、各カテゴリーからランダムに 1 個ずつデータを抽出し、その 2 個のデータのみで教師あり学習を行った結果、図 7.5 のように代表点従って識別境界がばらつき 10 回の試行の結果識別率の平均は 82.3% となった。このときの代表点 r_1, r_2 の共分散行列の行列式の値は 529051.7 であった。次に図 7.5 と同じ教師ありデータを用いて $\alpha = 1000$ で混合学習した結果図 7.6 のようになり、10 回の試行の結果識別率の平均は 100% となった。このときの代表点の共分散行列の行列式の値は 1.32886 であった。なお $\alpha = 1000$ という値は一見大きく思われるが、教師なしデータと教師ありデータの個数の比とデータ分布とを考慮すれば、教師あり学習の重みが教師なし学習よりも若干大きく設定される程度の値である。

次にもう少し高次元のデータの例として Olivetti の顔画像データベースから 2 人分のパターンを選んで識別を行った。各人のパターンは向きの異なるモノクロ 92×112 の顔画像である。このままではデータの次元が高すぎるので、Goudail ら [61] に倣い高次自己相関関数により 25 次元に落としたデータを用いた。代表点は 2 個とした。まず全データで教師なし学習をした結果図 7.7 のようになり、識別率は 75% となった。この図は 25 次元データを 3 次元に射影して示したものである。次に図 7.8 は全データで教師あり学習をした結果であり、95% の識別率となった。次に教師ありデータの数が少ない教師あり学習として、各カテゴリーからランダムに 2 個ずつデータを抽出し、その 4 個のデータのみで教師あり学習を行った結果、図 7.9 のように代表点がばらつき 10 回の試行の結果識別率の平均は 89% となった。次に図 7.9 と同じ教師ありデータを用いて $\alpha = 500$ で混合学習した結果図 7.10 のように代表点のばらつきが図 7.9 より

も小さくなり，10回の試行の結果識別率の平均は91.5%となった．以上のように教師ありデータの数が少ないときには，教師なしデータも加えて混合学習した方が識別率が向上することが確認された．

7.4 むすび

教師ありデータと教師なしデータが混在する場合の最近傍パターン識別器の学習法として勾配法に基づく方法を提案した．この方法では量子化誤差と識別誤差が加重係数により単調に変化することと，混合学習により代表点のばらつきが減少し，それによって識別率の期待値が増えることが期待されることを理論的に解析した．最後にこれらの理論的な予想を典型的な例による実験によって確認した．ここで扱ったのは最近傍識別器という最も基本的なパターン識別器であったが，ここでの結果はニューラルネットを含む更に高度なパターン識別器の混合学習の基礎となると思われる．

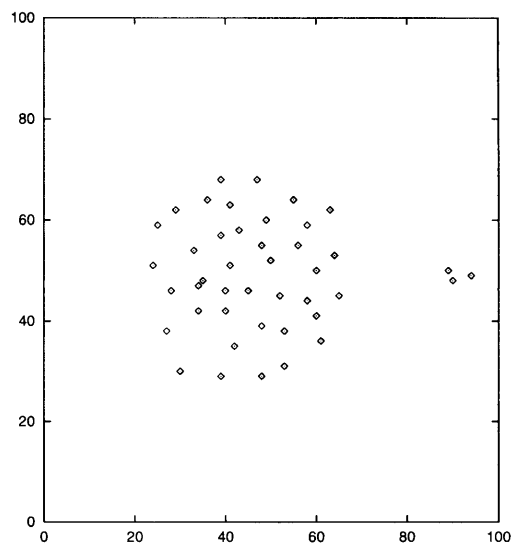


図 7.1: 学習データ

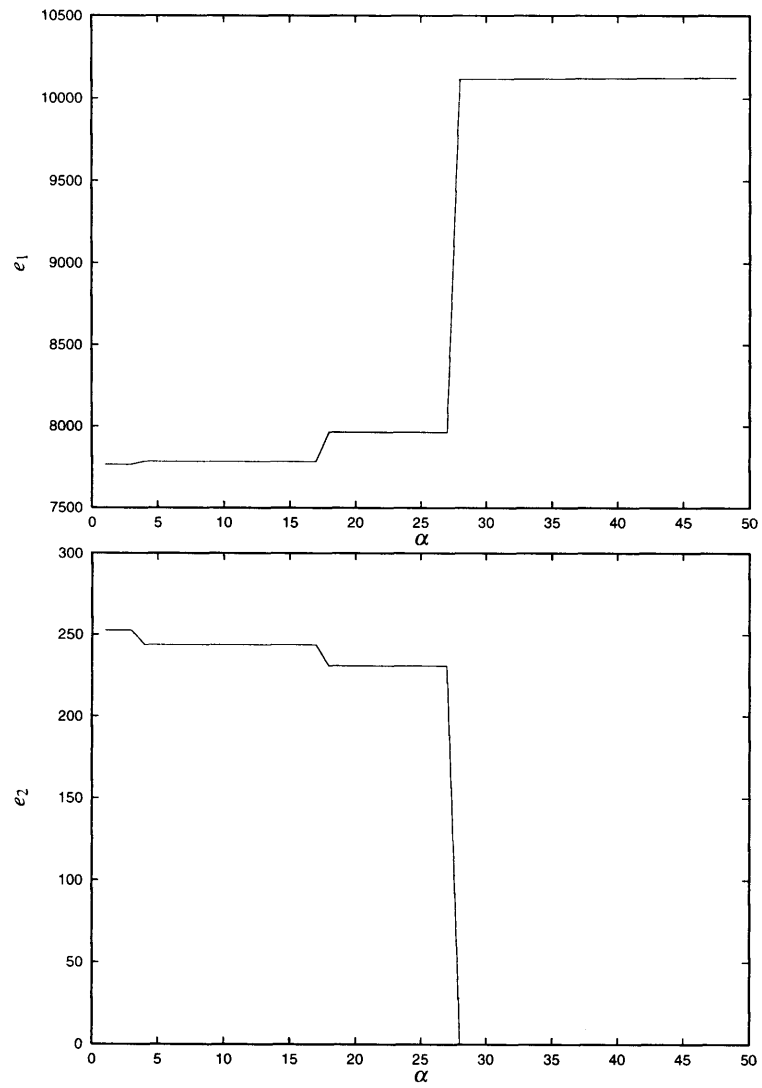


図 7.2: 量子化誤差 e_1 と識別誤差 e_2 の α による変化

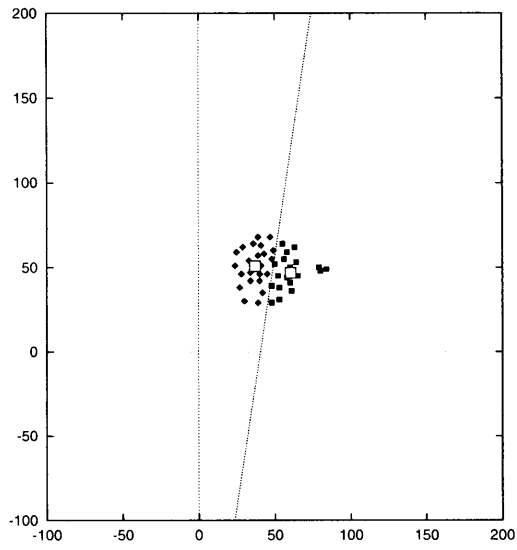


図 7.3: 教師なし学習の代表点 (□) と識別境界 (点線)

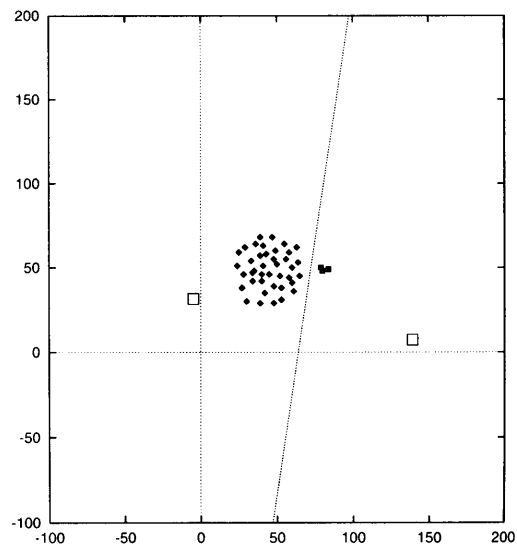


図 7.4: 全データによる教師あり学習

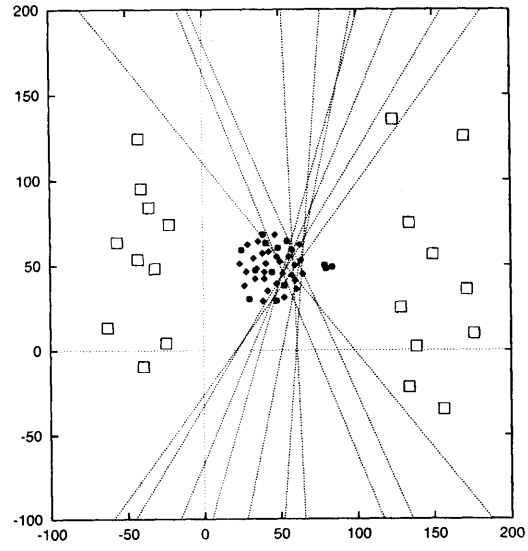


図 7.5: 2 個のデータによる教師あり学習

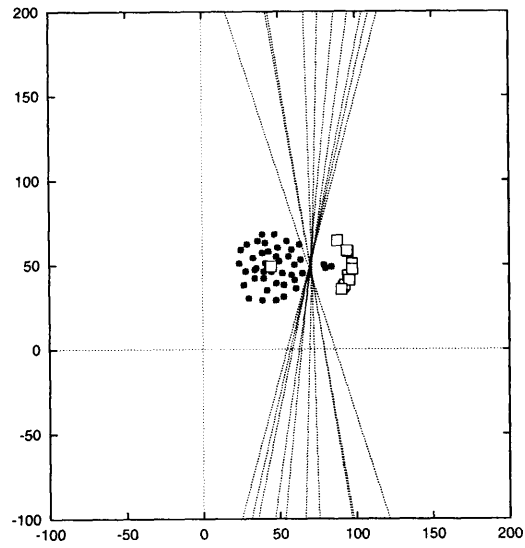


図 7.6: 混合学習の結果

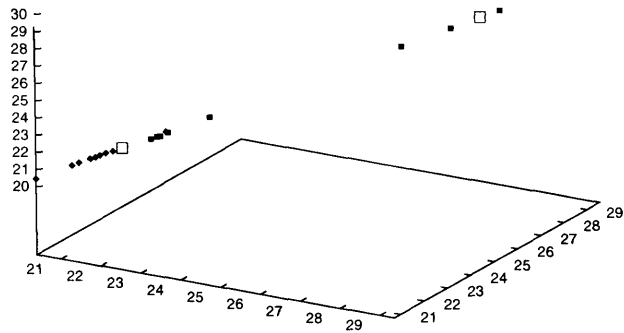


図 7.7: 顔画像データの教師なし学習

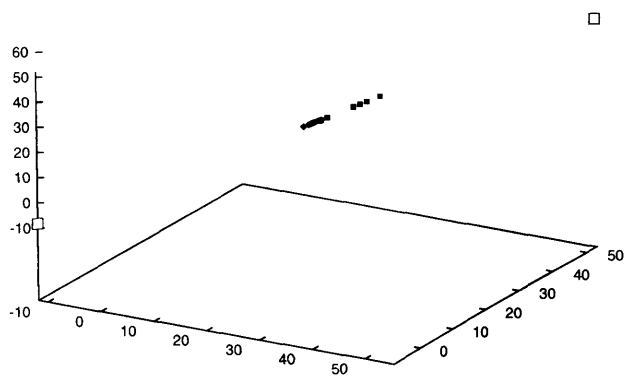


図 7.8: 顔画像データの教師あり学習

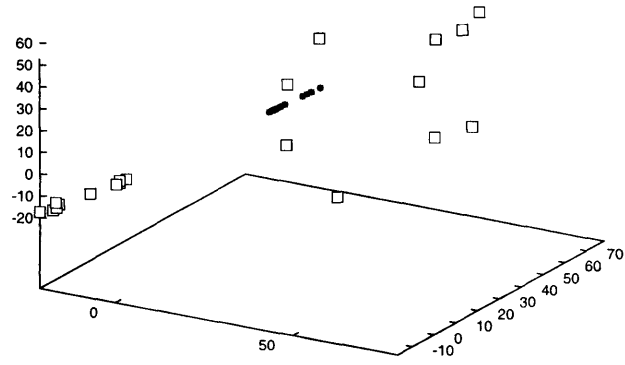


図 7.9: 4 個の顔画像データによる教師あり学習

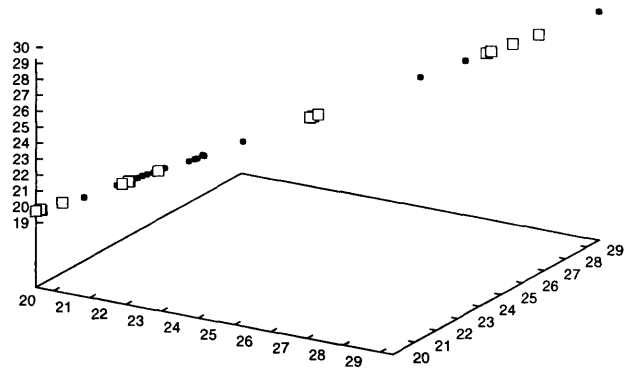


図 7.10: 顔画像データによる混合学習の結果

第 8 章

マルチモーダルパターン識別器の教師なし学習

8.1 まえがき

この章ではこれまでに述べた知覚と学習の問題を統合して，マルチモーダルパターンの認識モデルと学習法を提案する．画像や音声など複数の属性を合わせ持つマルチモーダル情報の統合処理としては，複数特徴量の測定によるパターン識別 [65]，複数センサーによる物理量推定 [66]，異種モード間の物理量変換 [67, 68] などがある [69]．本章では最初のタイプすなわちマルチモードの物理量が入力され，入力データの категорияが出力される場合を考える．このような場合の典型的な心理実験としてはマガーク (McGurk) 効果が知られている [70]．これは視覚入力と聴覚入力の category が異なる (例えば ga の唇の形と ba の音声) とき，そのどちらとも異なる category の音 (da) が知覚されるというものである．Massaro [71] はファジイモデルに基づく知覚モデル FLMP を提案し，マガーク効果を含む視聴覚実験結果を説明し，また FLMP がベイズ則による統合と等価であることを示している．このほかにも視聴覚統合のモデルが提案されている [72] が，いずれもノイズの影響が説明されておらず不十分であり，また学習が組み込めるようなモデルではない．ニューラルネットによるモデル化には学習が重要である．丹野ら [73] は教師あり学習のモデルによりマガーク効果を再現して

いるが教師なし学習の方が自然である。de Sa[74] や柴田ら [75] は信号入力同時性に基づいて2モード入力を統合する相互教師学習とも呼べる教師なし学習アルゴリズムを提案し、石井ら [76] は更に任意の数のセンサー情報を統合するニューラルネットワークシステムの教師なし学習方を提案している。これらのモデルでもマガーク効果などを説明できるであろうが、ノイズの影響など調べられておらず、またモデルがかなり複雑であるため動作の原理的な機構が理解し難い。赤穂ら [77] は次元圧縮を伴うEMアルゴリズムによる教師なし学習によりマルチモーダル情報源からの概念獲得をモデル化した。本章ではパターン認識器のなかで最も簡単な最近傍識別器を用い、統合法としてもベイズ則に基づく単純な確率の積による方法を考え、教師なし学習のアルゴリズムを導き、マガーク効果を含め、典型的な統合化の性質について調べ、多重画像のセグメンテーションへの応用例を示す。情報統合の様式に関して本論文の識別器が赤穂らのものと異なる点は、ロバスト推定によって多数決となっていることである。

8.2 最尤推定による識別と学習

データ d はマルチモーダルすなわち $d = [d_1, \dots, d_l]$ とする。例えば音声認識では d_1 が音声スペクトル、 d_2 が唇の形状などである。クラスが n 個 ($k = 1, \dots, n$) ある (例えばあいうえおの5個の母音) とする。 n は予め与えられるかあるいは適当な数に仮定するとする。テストデータ d が所属するクラスを最尤推定

$$\max_{k \in \{1, \dots, n\}} p(d|k) \quad (8.1)$$

で決定することにする。 $p(d|k)$ は第 k クラスからパターン d が生じる確率である。通常各モードは互いに独立である [71] ので (8.1) は

$$\max_{k \in \{1, \dots, n\}} \prod_{i=1}^l p(d_i|k) \quad (8.2)$$

となる。対数をとると (8.2) は

$$\max_{k \in \{1, \dots, n\}} \sum_{i=1}^l \ln p(d_i|k) \quad (8.3)$$

となる。

以上は $p(d|k)$ の関数形が与えられたもとでの識別則であるが、次に学習用データ d が (例えば母音の発声データがランダムに) m 個与えられたもとでこの関数 $p(d|k)$ のパラメータ r を求める学習法について考える。 m 個のデータも各々マルチモーダル $[d_{1j}, \dots, d_{lj}] (j = 1, \dots, m)$ である。パラメータ r も最尤推定すなわち r の尤度関数 $p([d_{11}, \dots, d_{l1}], \dots, [d_{1m}, \dots, d_{lm}]|r)$ を最大にする r

$$\max_r p([d_{11}, \dots, d_{l1}], \dots, [d_{1m}, \dots, d_{lm}]|r) \quad (8.4)$$

で求める。 m 個のデータも通常互いに独立であるので (8.4) は

$$\max_r \prod_{j=1}^m p([d_{1j}, \dots, d_{lj}]|r) \quad (8.5)$$

となる。普通各クラスは互いに排他的である、すなわち

$$p([d_{1j}, \dots, d_{lj}]|r) = \sum_{k=1}^n p([d_{1j}, \dots, d_{lj}]|k)p(k) \quad (8.6)$$

である。また各クラスの発生率 $p(k)$ は k によらず一定であると仮定する (式 (8.1) をベイズ則から導く過程で既にこの仮定は使っている) と (8.5) は

$$\max_r \prod_{j=1}^m \sum_{k=1}^n p([d_{1j}, \dots, d_{lj}]|k) \quad (8.7)$$

となる。各クラスの各モードは互いに独立であるとしたので (8.7) は

$$\max_r \prod_{j=1}^m \sum_{k=1}^n \prod_{i=1}^l p(d_{ij}|k) \quad (8.8)$$

となる。対数をとると (8.8) は

$$\max_r \sum_{j=1}^m \ln \sum_{k=1}^n \prod_{i=1}^l p(d_{ij}|k) \quad (8.9)$$

となる。この式の (局所) 最適解を反復法で求めるのが学習である。本章では反復法として EM アルゴリズムを使うことにする。

8.3 確率分布の例

ここでは $p(d_i|k)$ の関数形を

$$p(d_i|k) = \frac{1}{z_i} (\epsilon_i + s_i e^{-a_i \|d_i - r_{ik}\|^2}) \quad (8.10)$$

としてEMアルゴリズムを具体的に示す． z_i は規格化定数である．(8.10)の第1項 ϵ_i は一様分布を表すので，ある有界な領域以外では $p(d_i|k) = 0$ であるとする．(8.10)の形の分布はロバスト推定で用いられる ϵ 汚染正規分布[78]の1例であり， ϵ_i の一様分布は外れデータ(outlier)の分布であり，強度 s_i の正規分布が正常データ(inlier)を表す．ノイズの多い状況下ではこのようなロバストな分布を仮定する必要があると予想される．分布が(8.10)の場合，式(8.3)は

$$\max_{k \in \{1, \dots, n\}} \sum_{i=1}^l \ln(\epsilon_i + s_i e^{-a_i \|d_i - r_{ik}\|^2}) \quad (8.11)$$

となり，(8.9)は

$$\max_r \sum_{j=1}^m \ln \sum_{k=1}^n \prod_{i=1}^l (\epsilon_i + s_i e^{-a_i \|d_{ij} - r_{ik}\|^2}) \quad (8.12)$$

となる．簡単のためここでは ϵ_i, s_i, a_i は適度な値に固定することにし，正規分布の中心 r_{ik} だけを学習することにする． r_{ik} は第 k クラスの代表点 r_k の第 i モード値($i = 1, \dots, l; k = 1, \dots, n$)である．式(8.12)を r_{ik} で微分して0とおくと

$$\frac{\sum_{j=1}^m [\prod_{i' \neq i} (\epsilon_{i'} + s_{i'} e^{-a_{i'} \|d_{i'j} - r_{i'k}\|^2}) (r_{ik} - d_{ij}) e^{-a_i \|d_{ij} - r_{ik}\|^2}]}{\sum_{k=1}^n \prod_{i=1}^l (\epsilon_i + s_i e^{-a_i \|d_{ij} - r_{ik}\|^2})} = 0 \quad (8.13)$$

となるから r_{ik} の適当な初期値 $r_{ik}^{(0)}$ から出発して

$$\begin{aligned} r_{ik}^{(\nu+1)} &= \left\{ \frac{\sum_{j=1}^m [d_{ij} e^{-a_i \|d_{ij} - r_{ik}^{(\nu)}\|^2} \prod_{i' \neq i} (\epsilon_{i'} + s_{i'} e^{-a_{i'} \|d_{i'j} - r_{i'k}^{(\nu)}\|^2})]}{\sum_{k=1}^n \prod_{i=1}^l (\epsilon_i + s_i e^{-a_i \|d_{ij} - r_{ik}^{(\nu)}\|^2})} \right\} \\ & \quad / \left\{ \frac{\sum_{j=1}^m [e^{-a_i \|d_{ij} - r_{ik}^{(\nu)}\|^2} \prod_{i' \neq i} (\epsilon_{i'} + s_{i'} e^{-a_{i'} \|d_{i'j} - r_{i'k}^{(\nu)}\|^2})]}{\sum_{k=1}^n \prod_{i=1}^l (\epsilon_i + s_i e^{-a_i \|d_{ij} - r_{ik}^{(\nu)}\|^2})} \right\} \end{aligned} \quad (8.14)$$

で r_{ik} を反復更新し， $r_{ik}^{(\nu)}$ の収束値を解とする．式(8.14)は第 k クラスに所属する d_{ij} の平均値が r_{ik} であることを表している．この学習法がEMアルゴリズムに一致することは容易に示せる．

8.4 識別の統合則について

式(8.10)の ϵ_i が統合において果たす役割を考えるためにまず $\epsilon_i = 0$ の場合について考える．各クラスに代表点 r_k が1個あり，これへの距離でパターン d を最近傍識別

すると

$$\min_{k \in \{1, \dots, n\}} \|d - r_k\|^2 \quad (8.15)$$

となり, これは

$$\max_{k \in \{1, \dots, n\}} e^{-\|d - r_k\|^2} \quad (8.16)$$

と同値である.

$$\|d - r_k\|^2 = \sum_{i=1}^l a_i \|d_i - r_{ik}\|^2 \quad (8.17)$$

とすると (8.16) は

$$\max_{k \in \{1, \dots, n\}} \prod_{i=1}^l s_i e^{-a_i \|d_i - r_{ik}\|^2} \quad (8.18)$$

となる (式 (8.10) との比較のため s_i をかけた). 式 (8.18) は各モードでの最近傍識別の AND による統合であり, 統合した結果も最近傍識別である. クリस्पな (すなわち a_i が十分大きい) 最近傍識別の場合には, 全てのモードでの識別結果が一致したときに限り統合系での識別結果が出力され, それ以外ではクラスは不明すなわちリジェクトされる. a_i が通常値のときには式 (8.18) はファジイとなり, 原理的にはどれかのクラスに識別されるが, 式 (8.18) の積はどれもほとんど 0 になり, 実質的には識別不能に近い. これは式 (8.18) が基本的にはこのような全員一致による統合則であるからである.

以上が $\epsilon_i = 0$ の場合である. 次に $\epsilon_i \neq 0$ の場合を考える. 式 (8.16) の最近傍則は

$$\max_{k \in \{1, \dots, n\}} \epsilon + s e^{-\|d - r_k\|^2} \quad (8.19)$$

としてもよい. これは式 (8.10) と同じ形であり, 従って式 (8.10) の各モードは最近傍識別であり, 式 (8.11) は $\epsilon_i = 0$ の場合と同じく最近傍識別を AND で統合したものである. しかし今度の場合は全員一致による統合則とは異なる. 典型的な場合として $\epsilon_i = 1, s_i = e - 1$ のときを考えてみる. すると $\ln(\epsilon_i + s_i e^{-a_i \|d_i - r_{ik}\|^2})$ の値は, $d_i = r_{ik}$ の (すなわちデータが第 k クラスに属す) とき 1 となり, d_i が r_{ik} から遠い (すなわち第 k クラスには属さない) と 0 になる. 従って $\sum_{i=1}^l \ln(\epsilon_i + s_i e^{-a_i \|d_i - r_{ik}\|^2})$ は, データが第 k クラスに属すと判定されたモードの総数, すなわち第 k クラスが獲得した票の総和となる. 従って式 (8.11) はモードの多数決による統合則である. 実際には s_i はモー

ド毎に異なるので投票は重み付きである。このように $\epsilon_i \neq 0$ の場合は多数決となり、 $\epsilon_i = 0$ の場合の全員一致よりも緩やかな統合則となる。例えばあるモードのデータ値が異常値である場合には、全員一致の場合にはリジェクトされるが、多数決では正しく識別される。

以上の統合識別則をニューラルネットに似たシステムで実行するには、対数をとる前の式 (8.2) の形の方がよいであろう。 $p(d_i|k)$ が式 (8.10) のときの式 (8.2) は図 8.1 のようなネットワークで表せる。図中 r_{ik} と記したニューロンは r_{ik} を中心とする動径基底関数ニューロンで、その出力は $p_{ik} = \epsilon_i + s_i e^{-a_i \|d_i - r_{ik}\|^2}$ である。また Π と書いたニューロンは入力の積を出力するニューロンである。最終段の WTA は入力のなかの最大のものを探す WTA ニューラルネットである。 ϵ_i は自発放電のパルス頻度であり、 s_i は入力の強度を表す。生理学実験でマルチモーダルニューロンのパルス頻度は各モードだけが入力されたときのパルス頻度の積に近いことが報告されている [79]。また上記の学習則については異なるモードの入力が同じ物を表すという対応付けが必要であるが、これは通常仮定される [74] ように入力の同時性によるものとする。この学習則は入力データの分布の近似によってクラスタリングを行う教師なし学習であるが、特にモードが2つのときには近似的に互いの識別結果を教師とする相互教師学習になる。

8.5 学習例

前記の学習はクラスタリングの1種であるが、通常のクラスタリングにはないマルチモード特有の性質がある。それらの性質が典型的に現れるいくつかの例についてシミュレーションしてみた。以下パラメータは特記したもの以外の全てのシミュレーションで $\epsilon_i = 0.1, s_i = 1.0, a_i = 0.02$ とした。

まず最初に上記の学習が、各々のモード毎に別々にクラスタリングするのとは違うことを図 8.2 のようなデータで調べてみた。モードは2つで各々のモードは2次元、すなわち $d = [d_1, d_2]$ で図 8.2(a) が d_1 で (b) が d_2 である。 d_1, d_2 とも2次元 $d_i = [d_{ix}, d_{iy}]$ ($i = 1, 2$) である。データの分布を正確に示すには図 8.2(a) のどの点と (b) のどの点とで1つのデータなのかを表示する必要があるが煩雑になるので省略する。以下の結

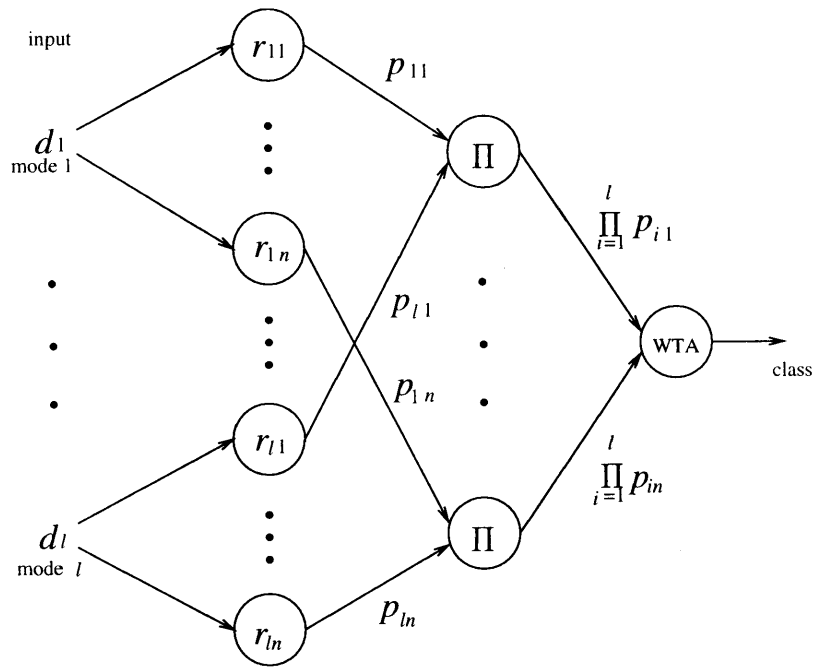


図 8.1: 式 (8.2) のネットワーク表示

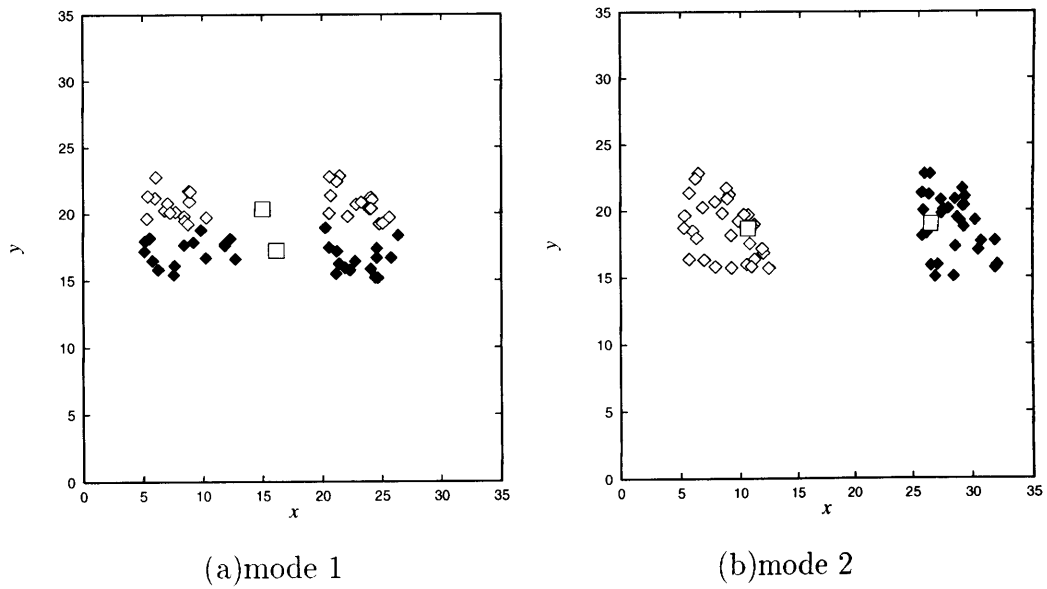
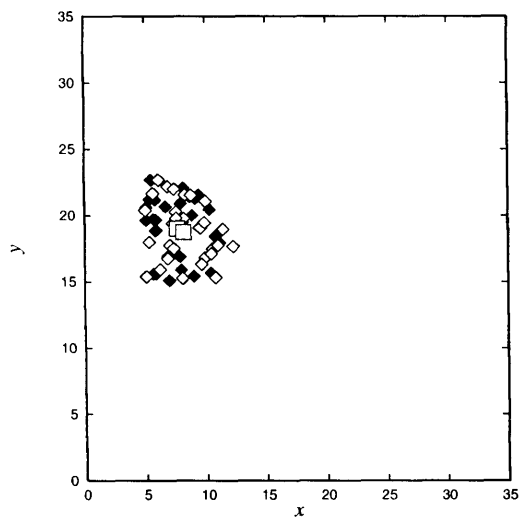
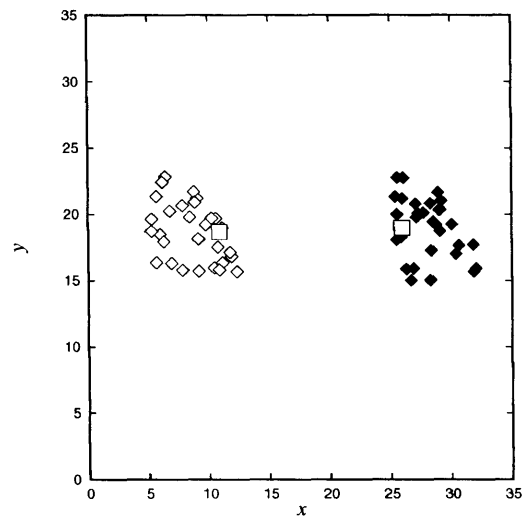


図 8.2: 実験例 1



(a)mode 1



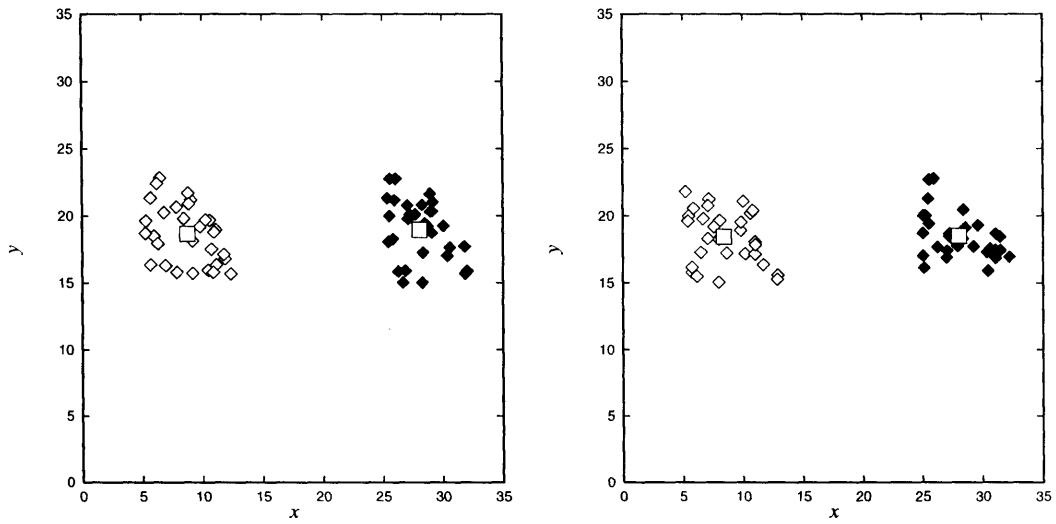
(b)mode 2

図 8.3: 実験例 2

果の理解には、図 8.2(a) の黒点と (b) の黒点とで、また図 8.2(a) の白点と (b) の白点とで 1 つのデータを構成するということだけで十分である。学習の結果、図中に少し大きめの白点で示す代表点 r_{ik} が求まった。グループ 1(白のデータ) の代表点のモード 1 での座標 r_{11} が図 8.2(a) 中の上の白四角、モード 2 での座標 r_{21} が図 (b) 中の左の白四角である。図 8.2(a),(b) 中のもう 1 つの白四角がグループ 2(黒のデータ) の代表点の各モードでの座標 r_{12} , r_{22} である。なおデータのこの黒と白へのグループ分けは予め教師情報として与えられたものではなく、学習で求まった代表点の配置による識別結果である。この結果は、モード 1 でのクラスタ間距離がモード 2 でのクラスタ間距離よりも小さいことから帰結されている。この関係が逆ならグループの別れ方もモード 1 と 2 とで逆になる。すなわちクラスタ分離が明瞭なモードがグループ分けを律する。この極端な場合が図 8.3 のように片方のモードでデータが 1 つのグループに融合してしまっている場合であり、当然ながらグループ分けはモード 2 だけで決まる。

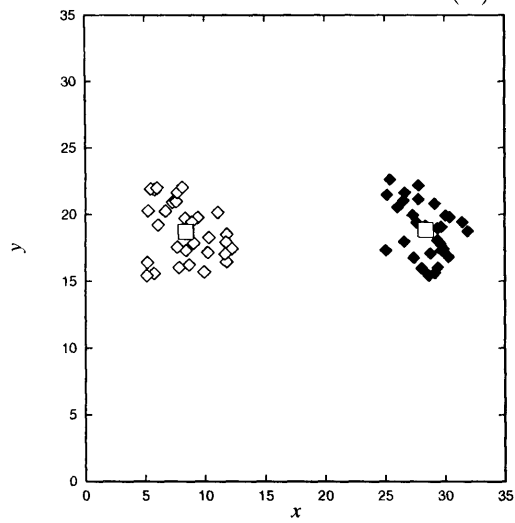
以上はモードが 2 つでも現れる性質であるが、次にモードが 3 つ以上で始めて現れる現象すなわち多数決の効果について図 8.4 のような 2 次元のモード 3 つの場合について調べてみた。図の表示は図 8.2 や 8.3 と同じである。図 8.5 は学習結果の $p(d_i|k)$ の値である。点線が $p(d_i|1)$ すなわちクラス 1(白のデータ) への帰属度、実線が $p(d_i|2)$ すなわちクラス 2(黒のデータ) への帰属度である。ただし正規化はしていない。横軸はデータの座標値ではなくデータの番号である。図 8.4 にはデータの番号を記入していないが、30 番までが白のデータで、31 番から 60 番までが黒のデータである。白のデータでは全てのモードで $p(d_i|1) = \epsilon_i + s_i > p(d_i|2) = \epsilon_i$ となっている(黒ではこの逆)。図 8.6 は図 8.5 の 3 つの積であり、統合結果での識別は各モードでの識別と一致する。

次に図 8.4 のデータにおいて、ある黒のデータのモード 1 での座標を白のデータのグループの中に移動してみる。また別の黒のデータについてモード 2 での座標を白のデータのグループの中に移動する。白のデータについても同様にあるモードで座標を黒へ移動させる。このようにして作ったのが図 8.7 のデータである。このようなデータで学習した結果、 $p(d_i|k)$ は図 8.8 のようになった。図 8.8 の横軸は図 8.5 と同じくデータの番号である。点線が第 1 グループ(白) への帰属度 $p(d_i|1)$ 、実線が第 2 グループ(黒) への帰属度 $p(d_i|2)$ である。例えば 13 番のデータは白のデータであるが、モード 1 で



(a)mode 1

(b)mode 2



(c)mode 3

図 8.4: 実験例 3

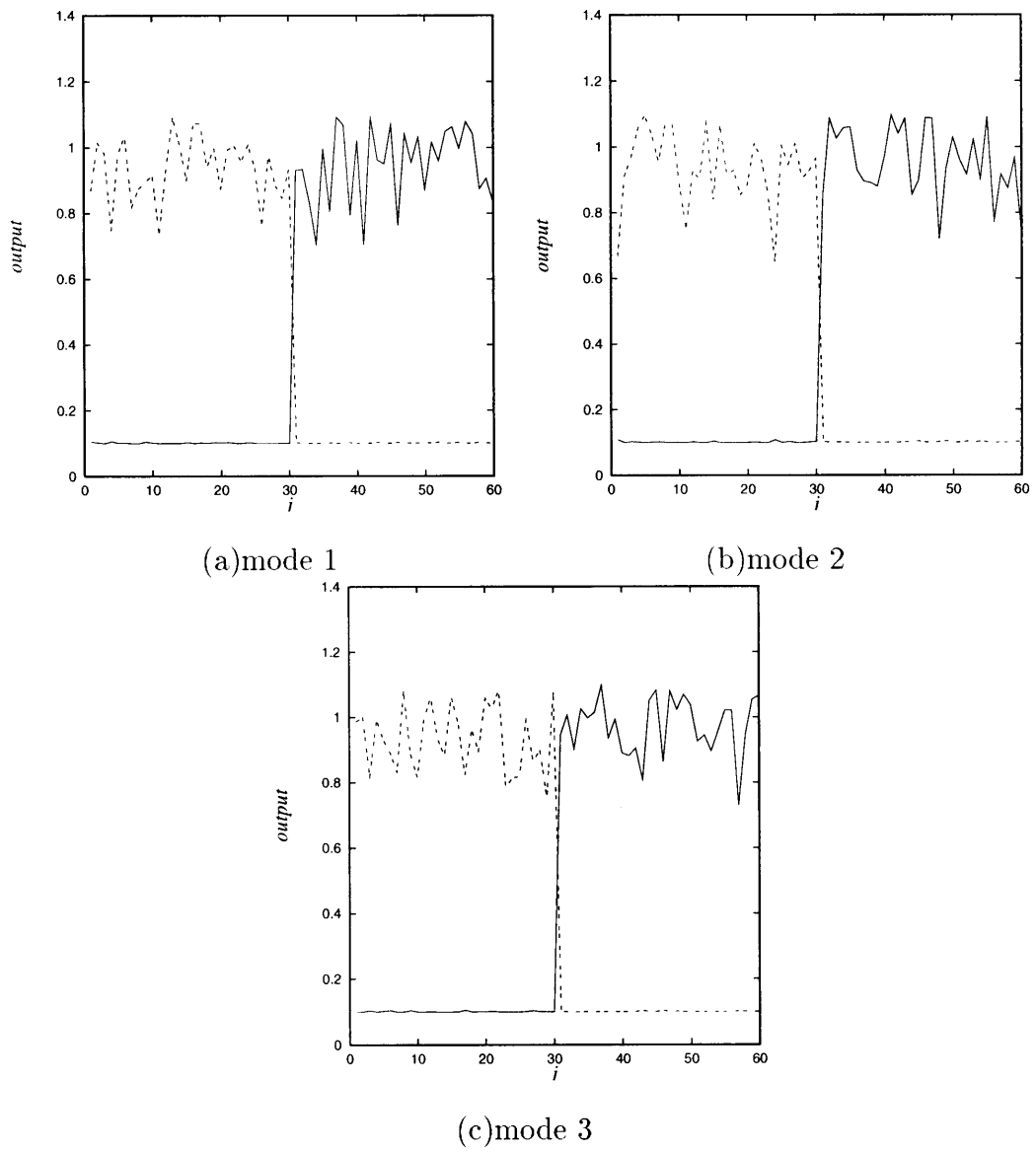


図 8.5: 図 8.4 での $p(d_i|k)$ の値

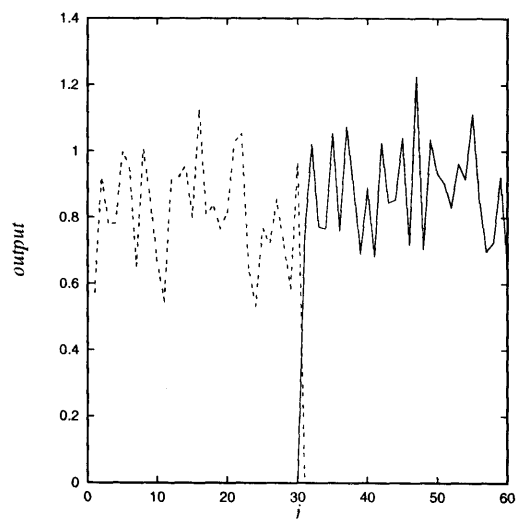
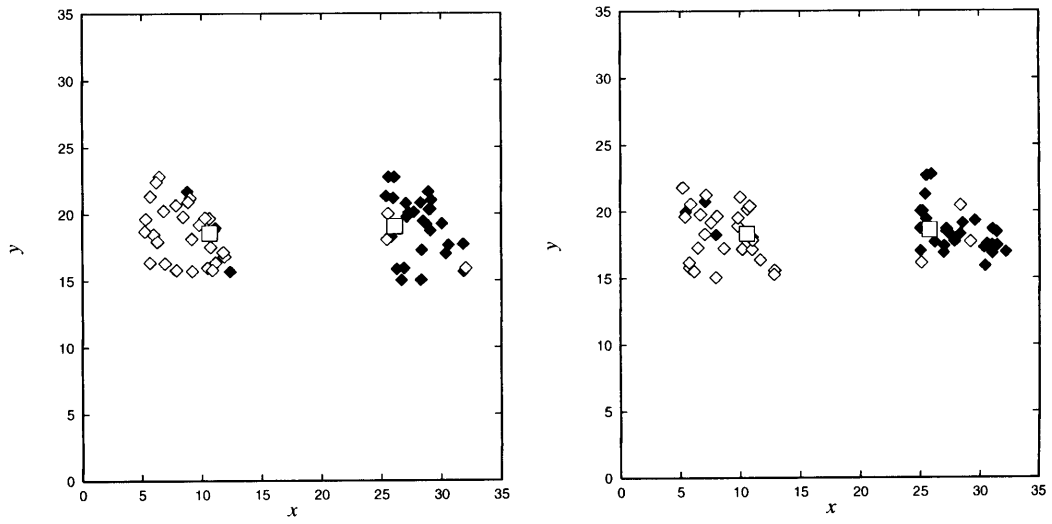
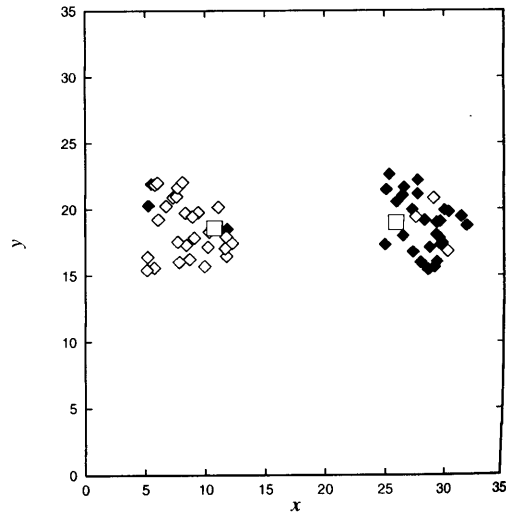


図 8.6: 図 8.5 の 3 つの $p(d_i|k)$ の積



(a)mode 1

(b)mode 2



(c)mode 3

図 8.7: 図 8.4 のデータをいくつか入れ換えたもの

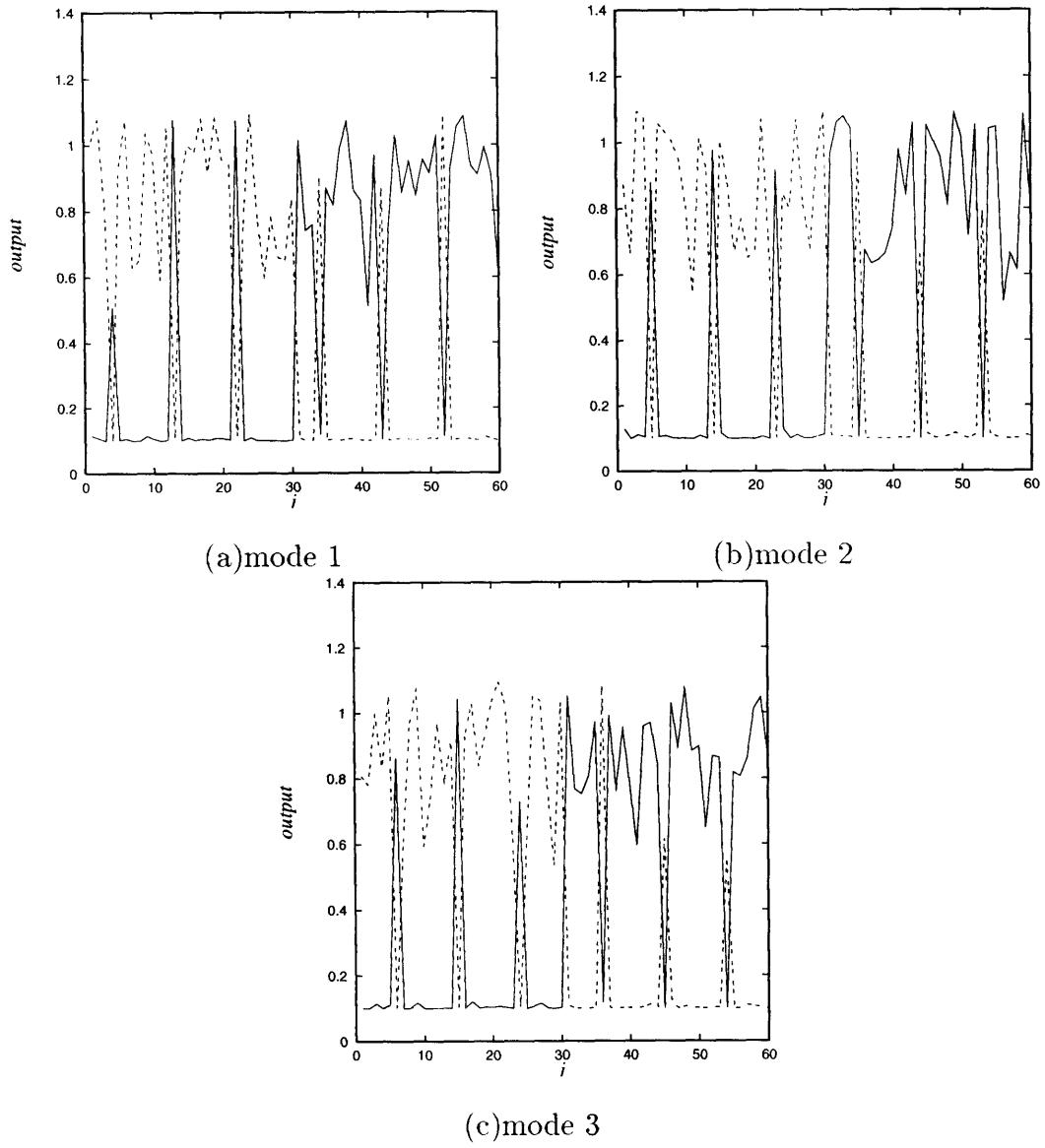


図 8.8: 図 8.7 での $p(d_i|k)$ の値

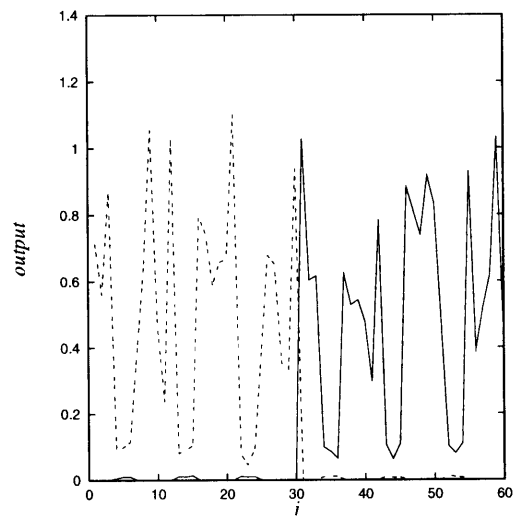
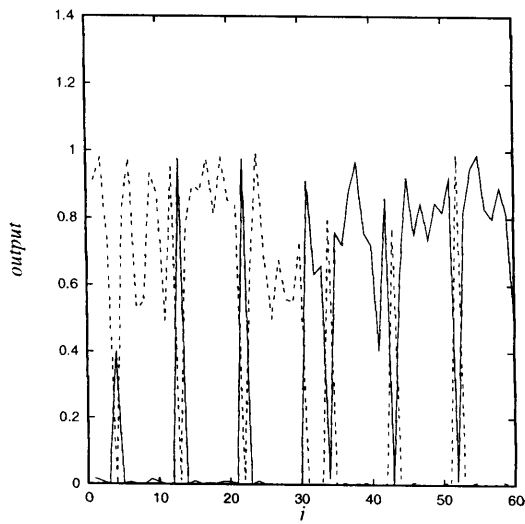
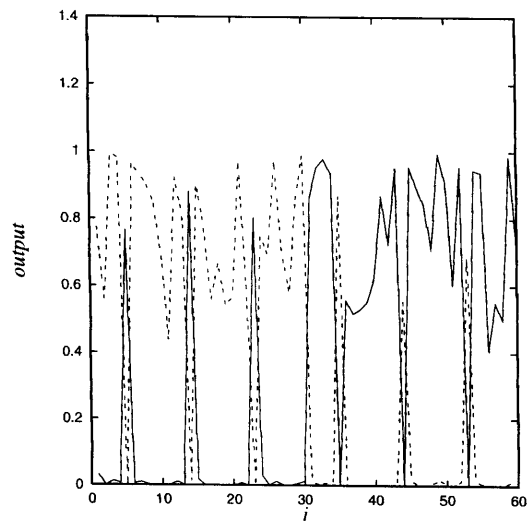


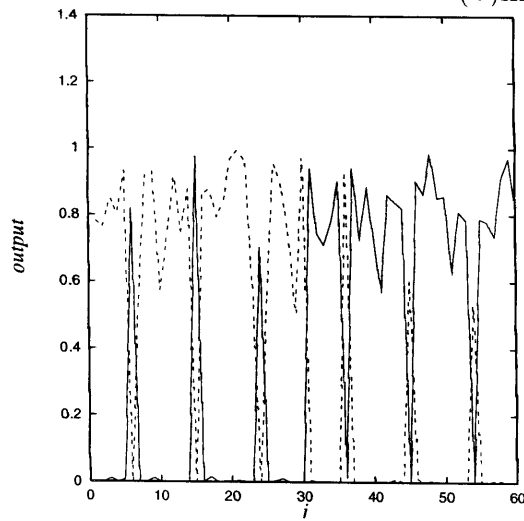
図 8.9: 図 8.8 での 3 つの $p(d_i|k)$ の積



(a)mode 1



(b)mode 2



(c)mode 3

図 8.10: $\epsilon = 0$ のときの $p(d_i|k)$

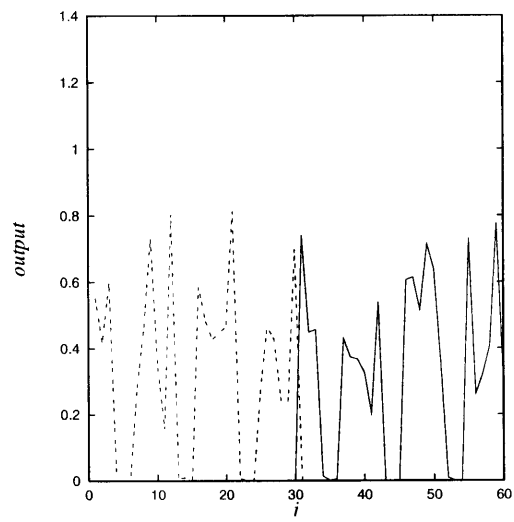


図 8.11: 図 8.10 での 3 つの $p(d_i|k)$ の積

の座標が黒の所に移動している。従って $p(d_1|1) = \epsilon_1 < p(d_1|2) = \epsilon_1 + s_1$ となり、図 8.5 のときと逆転している。しかしモード 2 と 3 では図 8.5 と同じであり $p(d_i|1) = \epsilon_i + s_i > p(d_i|2) = \epsilon_i (i = 2, 3)$ となる。従って $\prod_{i=1}^3 p(d_i|1) = \epsilon_1(\epsilon_2 + s_2)(\epsilon_3 + s_3) > \prod_{i=1}^3 p(d_i|2) = (\epsilon_1 + s_1)\epsilon_2\epsilon_3$ となり、統合識別は図 8.5 と同じになる。他のデータについても同様である。図 8.9 は図 8.8 の 3 つの $p(d_i|k)$ を書けたものであり、このように 1 つのモードでグループの混同が起っても多数決によって、図 8.4 のときと同じく白のデータが第 1 グループ、黒が第 2 グループと識別される。全員一致による統合ではこのようなロバストな識別ができないことを確認するために $\epsilon_i = 0 (i = 1, 2, 3)$ として図 8.7 のデータについて学習した結果 $p(d_i|k)$ は図 8.10 のようになり、これらの積は図 8.11 のようになった。この場合、座標を移動したデータでは $p(d_i|1)$ も $p(d_i|2)$ もともに 0 になっており、どちらのグループに属すのか決定できない。従って $\epsilon_i = 0$ だと識別則が全員一致となり、1 つのモードでも矛盾していると識別不能となる。

最後に視聴覚の統合認識において有名なマガーク効果について非常に簡単なモデル実験で調べてみた。これまでにいくつかのモデルでマガーク効果が説明されているが、学習が含まれていなかったり [71]、含まれていても教師あり学習であったりした [73]。またマガーク効果の重要な実験事実すなわち音声のみが明瞭なときにはマガーク効果は現れず、音声にノイズが加わったときにだけ現れる [80] という性質も説明されていない。ここではこのようなマガーク効果の性質を説明するために最も単純化したデータとして図 8.12 のようなデータでシミュレーションしてみた。聴覚を 1 次元空間で表しモード 1 として横軸とし、視覚も 1 次元空間で表しモード 2 として縦軸とする。クラスは典型例である “ga”, “da”, “ba” の 3 つとする。このときマガークら [70] の推測によればこれらのデータは図 8.12 のような布置になる。左下の黒点が “ga” であり、中央付近の黒点が “da” で、右上の黒点が “ba” である。実際にはデータはこれらの点の周りに分布すると考えられるが、マガーク効果を説明するにはこのような単純なデータで十分である。 $\epsilon_1 = 1, a_1 = 1, s_1 = 8, \epsilon_2 = 1, a_2 = 0.1, s_2 = 10$ としてこの 3 個のデータで学習した。 a_1 と a_2 の違いは、視覚よりも聴覚の方が信号の分散が大きい、すなわち単独モードの場合聴覚の方が視覚よりも各音を明瞭に識別できることを表す。これは心理実験とも一致する。また $\epsilon_1 = \epsilon_2 = 1$ という値はノイズが多い状況に対応するも

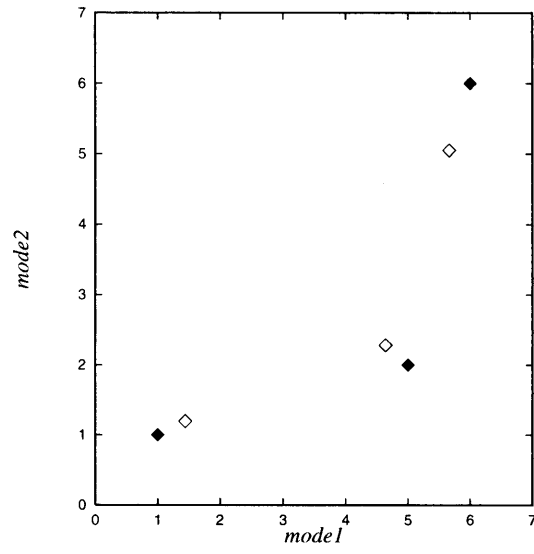


図 8.12: マガーク効果の実験例

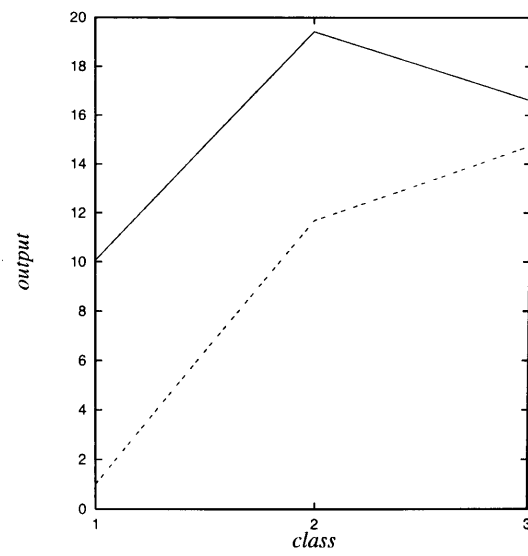


図 8.13: 出力 $\prod_{i=1}^2 p(d_i|k)$ の値

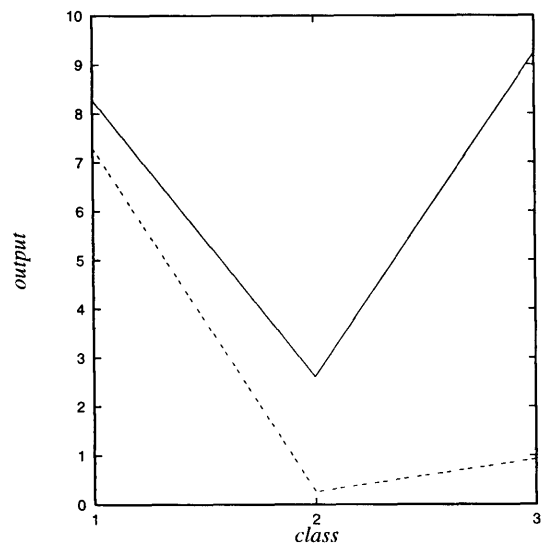


図 8.14: 視聴覚入力が逆のときの出力の値

のとする。学習の結果図 8.12の少し大きめの白点で記した代表点が得られた。この代表点で3つのクラスは正しく識別される。すなわち聴覚も視覚もノイズが多い状況のなかで学習が行われた結果そのような雑音下でも正しく識別されるようになった。当然 ϵ_1 や ϵ_2 が1より小さいすなわちノイズが学習時よりも小さい状況でも正しく識別できる。問題のマガーク効果ではモード1では“ba”の値が、モード2では“ga”が入力される。このときの $\prod_{i=1}^2 p(d_i|k)$ の値は図 8.13のようになる。横軸は k である ($k = 1$ が“ga”, $k = 2$ が“da”, $k = 3$ が“ba”である)。実線は $\epsilon_1 = 1, \epsilon_2 = 0.1$ のときすなわち聴覚にノイズがあり、視覚にはあまりノイズがないときであり、点線は $\epsilon_1 = 0.1, \epsilon_2 = 0.1$ のときすなわち聴覚、視覚ともノイズが少ないときである。前者の場合にはマガーク効果が起って視覚と聴覚の中間の“da”が認識され、後者ではマガーク効果は生じず聴覚入力である“ba”が知覚される。このように以上の単純なモデルで心理実験を定性的に再現できる。なおマガーク効果のときとは逆に視覚に“ba”, 聴覚に“ga”を入力した場合には図 8.14のようになり、ノイズが小さいときには聴覚入力の“ga”が知覚されるが、聴覚のノイズが大きいと“ba”と“ga”の出力が同じくらいに大きくこれらの音のどちらかあるいは2つの結合音が認識されると予想される。これも心理実験 [70] と一致する。

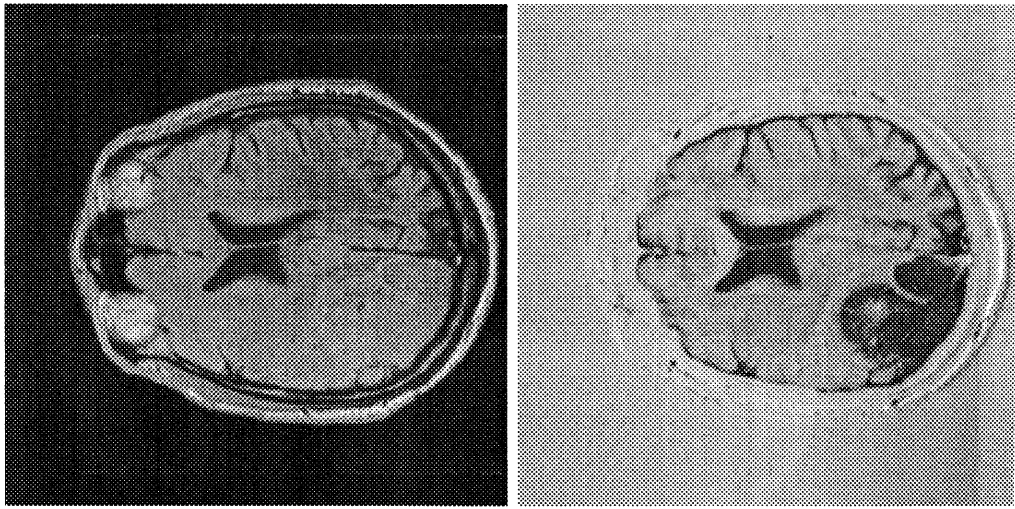
8.6 多重画像のセグメンテーション

医用画像や衛星画像などの多重画像はマルチモーダルデータの1例である。ただし画像ではモードは互いに独立であるが、画素は互いに独立ではない。画素間の依存性はマルコフランダム場などで扱われるが、最も粗い近似のセグメンテーションでは画素は独立に扱われる。ここでも画素は互いに独立と仮定して、上記の方法を多重画像のセグメンテーションに応用してみた。まず図 8.15に示す、プロトン密度とスピン緩和時間の異なる2枚の脳のMRI画像に適用してみた。画像 a(モード1)では右下の腫瘍が不鮮明であり、一方画像 b(モード2)では眼が不鮮明である。すなわち図 8.3と同じような状況になっている。図 8.16は本方法によって10個の領域に分割した結果であり、腫瘍も眼もセグメントとして切り出されている。次にSIDBAのカラー(RGB)

画像 “house” の成分画像 house-R(モード 1), house-G(モード 2), house-B(モード 3) それぞれの画像においていくつかの画素を互いに交換して図 8.7 と同じような状況のデータを作ったものが図 8.17 である。このデータ画像を 6 つの領域に分割した結果が図 8.18 である。多数決の効果によりノイズがほとんど取り除かれて正常なセグメンテーションが得られている。

8.7 むすび

ロバスト推定とベイズ則に基づくマルチモーダル情報の統合による識別法を提案して、それが多数決をファジイ化したものであることを示し、EM アルゴリズムによる教師なし学習法を導いた。この識別法の基本的な性質を単純なデータで検証し、マガーク効果を定性的に説明した。また多重画像のセグメンテーションへの応用例も示した。ここではクラスの代表点だけを学習させたが、より一般に全てのパラメータを学習で求める実験をするのが今後の課題である。またここでは統合による識別について調べたが、複数センサの測定値の統合や異種モード間の物理量変換について調べるのも今後の課題である。



(a)mode 1

(b)mode 2

図 8.15: 脳の MRI 画像

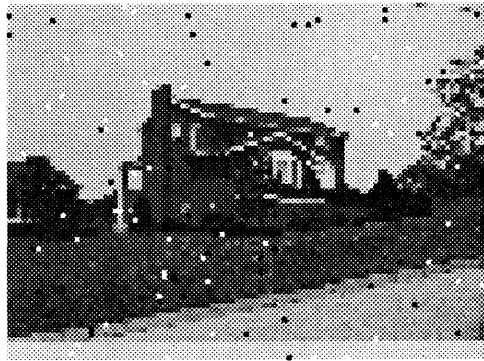


図 8.16: 図 8.15 の画像のセグメンテーションの結果



(a)mode 1

(b)mode 2



(c)mode 3

図 8.17: R,G,B それぞれにノイズを加えた house 画像

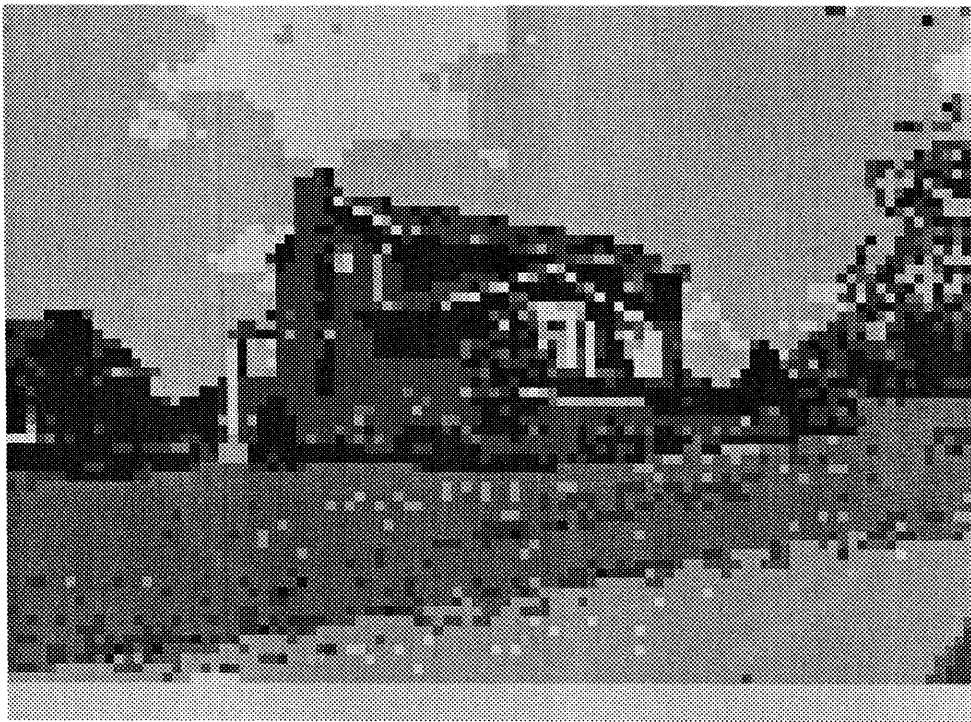


図 8.18: 図 8.17 の画像のセグメンテーションの結果

第 9 章

結論

本論文では、視覚パターンの知覚や学習の混合整数計画問題による数理モデル化について研究した結果をまとめた。

本研究によって得られた研究成果の概要は以下のとおりである。

まず第 1 章では、本研究の背景と扱っている問題を示し、あわせて論文の概要について述べた。

次に第 2 章では視覚パターンの知覚や学習を混合整数計画問題によってモデル化する際に必要となる数学的基礎を概括し技術的な問題を論じた。ここでは、生体の情報処理過程をパターン認識システムとして捉え、認識システムの出力が最適化問題の最適解として与えられると仮定した。認識システムの出力はカテゴリーのように整数の時もあれば物理量のように実数の時もあるので、認識システムは混合整数計画問題として定式化できることを述べた。また、認識システムの学習についても混合整数計画問題として扱えることを示した。これら認識や学習の定式化を最近傍識別を例にとり具体的に説明を行なった。更に認識システムとしての混合整数計画問題の解法として勾配系を用いて解くアナログ解法を提案した。このアナログ解法は生体のニューラルネットのモデルとして捉えることができる。

第 3 章ではラインプロセスを拡張した結合プロセスを提案し、画像処理への応用を示した。ラインプロセスではしきい値が絶対的な値に固定されるのでグレイレベルのスケール変化によって結果が変わるという欠点がある。結合プロセスはこの欠点を緩

和しエッジを保存して平滑化を行なうことができることを計算機実験により示した。またインパルスノイズの除去もできるように拡張し、この方法がガウスノイズの除去においてはメディアンフィルタより優れているという結果が得られた。結合プロセスを用いるとスパースなデータからの復元やセグメンテーションを容易に行なえる。更に結合プロセスに与えるパラメータを変化させたり、非線形拡散方程式を用いることによりマルチスケール表現が得られる。これらの処理について計算機実験を行ない結合プロセスの有効性を確認した。

第4章では結合プロセスが生体の視覚パターン知覚のモデルとしても有効であることを示すために、結合プロセスを用いた視覚パターン知覚のモデルを提案し幾何学的錯視の計算機実験を行なった。パターンの知覚体制化には近接性、類同性、連続性、対象性、閉合の法則があるが、本論文ではドットパターンを対象として最も基本的であると考えられる近接要因による群化のモデルを提案した。このモデルを作成するにあたって前提とした心理実験結果の性質は非線形性とスケール不変性である。Muller-Lyer 図形を対象として計算機実験を行ない変位量の形や不動点の位置が心理実験の結果を再現していることを確認した。また、矢線の角度や長さの変化による錯視量の変化についても心理実験の結果と定性的に一致した結果が得られた。更に同心円錯視、分割距離錯視、曲率の錯視や代表的な幾何学的錯視図形に対して計算機実験を行ない錯視現象のモデルとしての結合プロセスの有効性を確認した。しかし、ツェルナー、ポッケンドルフ、ヘリング、ブント図形についてはこのモデルで錯視現象を再現することができなかった。これらの錯視を再現するためには、体制化の他の要因を組み入れる等のモデルの拡張が必要であると考えられる。

第5章ではクラスタリング関数回帰による最近傍識別器の教師あり学習法を提案した。この方法は識別率が最小となるように最適化問題として定式化して教師信号に基づいて定数関数への回帰を学習することによって、最近傍識別器の代表点の最適な配置を決める方法である。この学習法を用いると代表点の数を十分多く準備しておけば、各カテゴリーに必要最小限の数の代表点が割り当てられ、不要な代表点はどのカテゴリーにも割り当てられずに余ることが計算機実験によって示された。またこの学習法を木探索識別器の代表点を一括して最適化する学習が行なえるように拡張した。二段

木探索クラスタリングにおいて回帰する関数の次元が学習の効率に与える影響を調べ、次元数が低い場合には一括学習の方が逐次学習より良い結果が得られることを計算機実験により確認した。

第6章では第5章で提案した方法の特徴抽出すなわちデータベクトルの次元の圧縮を行なうように拡張しその有効性を実験により示した。次元の圧縮には最も基本的であるアフィン部分空間への線形射影を用いた。この場合カテゴリーの代表が点ではなく超平面で表され、入力データを平面へ射影した残差が最も小さいカテゴリーに分類する。この超平面を決定するために、教師なし学習では射影したデータの復元誤差が最小となるように最適化問題として定式化した。教師あり学習では第5章で提案したクラスタリング関数回帰による方法を用いた。部分空間の取り方として、射影行列がクラス毎に異なり原点は同じ、射影行列が同じで原点がクラス毎に異なる、射影行列も原点もクラス毎に異なる、の3通りを考えそれぞれに対して簡単なデータを用いて実験することによりその特性を明確にした。更にこの方法によりパラメータ化された顔データの分類を行ない高い識別率が得られた。射影行列は一種の計量行列であり、カテゴリーとの距離を測るには他に様々な計量行列が考えられる。本論文で扱った方法は最も基礎的なものであり、人間のパターンに対する心理的な類似度に近い計量を求めることは今後の課題である。

第7章では教師ありのデータと教師なしのデータが混在する場合の最近傍識別器の学習法を提案した。ここでは、代表点の配置を決定するのに識別誤差と量子化誤差を共に最小化するように最適化問題として定式化した。この方法では荷重係数を変化させると量子化誤差と識別誤差が単調に変化することを理論的に示した。更に教師ありデータだけで学習するより、教師なしデータも付け加えて学習する方が識別率の期待値が向上することを、理論的な解析で予想した。簡単なデータと顔画像データに対し高次自己相関関数により25次元に落したデータを用いた計算機実験によってこの予想を実証した。

第8章ではロバスト推定とベイズ則に基づく複数の入力からなる異種データの統合すなわちマルチモーダル情報による最近傍識別法を提案した。この方法は多数決をフジイ化したものであることを示し、EMアルゴリズムによる教師なし学習法を導出し

た。この識別法の基本的な性質を簡単なデータで検証し、視覚と聴覚の統合の例としてマガーク効果を定性的に説明した。更に多重画像のセグメンテーションの計算機実験を行い応用例を示した。今後の課題として複数センサの測定値の統合や異種モード間の物理量変換についての研究がある。

以上本論文で提案したモデルはいずれも基本的な性質や機能に絞って抽象化したものであり、実際の生体の情報処理の振る舞いとは定性的に異なるところが少くない。更に精緻なモデルとするためには大規模なデータを用いて計算機実験を行ない、心理実験によって得られたデータとの比較検討によりモデルを拡張する必要がある。一方、モデルの精緻化や個別化とは逆に生体の認知の特質を示す一般的、統一的なモデルも興味のあるところであり、今後に残された課題である。これら生体の知覚や学習のモデルには、個人識別、文字認識、画像処理など幅広い様々な応用分野がある。しかしここで述べたモデルは生体での処理を意識したアナログ解法を用いているので計算機での実行は処理に要する時間が非常に大きい。実用化に際しては計算機で処理するのに適した高速な解法や専用のハードウェアの開発が望まれる。

謝辞

本研究を遂行し、また、まとめるにあたり、多くの先生方に御指導、御助言及び御鞭撻を頂きました。深く感謝いたします。

九州芸術工科大学 芸術工学部 画像設計学科 浦浜喜一教授には、本研究を進めるにあたり、終始御指導いただきました。本研究の核となる部分は、浦浜教授の熱心な議論と指導なくしては、成り立ち得なかったものです。

九州芸術工科大学 芸術工学部 画像設計学科 瀧山龍三教授、長島健次教授、福島重廣教授には、有効適切な御意見と御助言を賜りました。又、坂本博康講師、小野直樹助手には、有益な御助言及びはげましを頂きました。

株式会社富士通九州システムエンジニアリング 柴田善次郎社長、富永英俊取締役には、本研究の機会を与えて下さり、有益な御助言を頂きました。又、関連部署の方々には、円滑に研究が進められるよう御協力及びあたたかいはげましを頂きました。

以上、御指導、御協力頂いた方々に対し、あらためて深謝いたします。

参考文献

- (1) Metzger W. :“Gesetze des Sehens,” Kramer, 1953.
- (2) Kanizsa G. :“Organization in vision : Essays on Gestalt perception,” Praeger, 1979.
- (3) Grossberg S. and Mingolla E. : “Neural dynamics of form perception ; Boundary completion, illusory figures, and neon color spreading,” *Psychological Review*, 92, pp.173-211, 1985.
- (4) Grossberg S. and Mingolla E. : “Neural dynamics of perceptual grouping ; Textures, boundaries, and emergent segmentation,” *Perception and Psychophysics*, 38, pp.141-171, 1985.
- (5) Medin D.L. and Schaffer M.M. : “Context theory of classification learning,” *Psychological Review*, 92, pp.286-316, 1978.
- (6) Posner M.I. and Keele S.W. : “On the genesis of abstract ideas,” *Journal of Experimental Psychology*, 77, pp.353-363, 1968.
- (7) Reed S.K. : “Pattern recognition and categorization,” *Cognitive Psychology*, 3, pp.382-407, 1972.

- (8) von der Heydt R., Peterhans E., and Baumgartner G. : "Illusory contours and cortical neuron responses," *Science*, 224, pp.1260-1262, 1984.
- (9) 今野浩 : "整数計画法," 産業図書, 1981.
- (10) 福島雅夫 : "非線形最適化の理論," 産業図書, 1980.
- (11) Poggio T. and Girosi F. : "Networks for approximation and learning," *Proceedings of the IEEE*, 78, pp.1481-1497, 1990.
- (12) Li S.Z.: "Markov random field modeling in computer vision," Springer-Verlag, 1995.
- (13) Hopfield J.J. : "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences U.S.A.*, pp.2554-2558, 1982.
- (14) Danzig G.B., Fulkerson D.R. and Johnson S.M. : "Solution of a large scale traveling salesman problem," *Operations Research* 2, pp.393-410, 1954.
- (15) Bellman R.E. : "Dynamic Programming," Princeton University Press, 1957.
- (16) Holland J.H. : "Adaptation in Natural and Artificial Systems," University of Michigan Press, 1975.
- (17) Boucart M. and Delord S.: "The computation of contour information in complex objects," *Perception*, 23, pp.399-409, 1994.
- (18) Field D.J., Hayes A. and Hess R.F.: "Contour integration by the human visual system: evidence for a local associative field," *Vision Res.*, 33, pp.173-193, 1993.
images," *IEEE Trans. Patt. Anal. Mach. Intelli.*, 6, pp.721-741, 1984.
- (19) Geman S. and Geman D.: "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of

- (20) Kullback S. : "Information theory and statistics," John Wiley & Sons, 1959.
- (21) Kirkpatrick S., Gellat C.D. and Vecchi M.P. : "Optimization by simulated annealing," Science, 220, pp.671-680, 1983.
- (22) Rose K., Gurewitz E., Fox G. : "A deterministic annealing approach to clustering," Patt. Recog. Lett., 11, pp.589-594, 1990.
- (23) 乾敏郎: "Q&A でわかる脳と視覚," サイエンス社, 1993.
- (24) Koch C., Marroquin J. and Yuille A: "Analog neuronal networks in early vision," Proc. Nat. Acad. Sci. USA, 83, pp.4263-4267, 1986.
- (25) Lumsdaine A., Wyatt J. and Elfadel I.: "Nonlinear analog networks for image smoothing and segmentation," J.VLSI Signal Process., 3, pp.53-68, 1991.
- (26) 浦浜喜一, 上野真一郎, 鈴木道俊: "画像の多重スケール復元のアナログ解法," 信学技報, PRU92-105, 1993.
- (27) Perona P. and Malik J: "Scale-space and edge detection using anisotropic diffusion," IEEE Trans. Patt. Anal. Mach. Intelli., 12, pp.629-639 ,1990.
- (28) Watt R.J. : "Visual processing: computational psychological and cognitive research," Lawrence Erlbaum Associates 1988.
- (29) 横瀬善正: "視覚の心理学," 共立出版, 1956.
- (30) 中島和樹, 菅沼義昇, 伊藤正美: "心理的ポテンシャル場とニューラルネットによる図形認識," 信学技報, PRU90-155, 1990.
- (31) 関和徳, 杉江昇, 杉原厚吉: "図形変換規則を用いた錯視現象のモデル化," 信学論, J70-D, pp.2668-2677, 1987.
- (32) Zahn C.T.: "Graph theoretical methods for detecting and describing Gestalt clusters," IEEE Trans. Comput., C-20, pp.68-86, 1971.

- (33) 和田陽平, 大山正, 今井省吾: "感覚知覚心理学ハンドブック," 誠信書房, 1969.
- (34) Stevens K.A.: "Computation of locally parallel structure," *Biol. Cybern.*, 29, pp.19-28, 1978.
- (35) Kohonen T.: "Self-organization and associative memory, 3rd ed.," Springer, 1989.
- (36) N. Kambhatla and T.K. Leen, "Fast nonlinear dimension reduction," *NIPS*, 6, pp.152-159, 1994.
- (37) T. Kohonen, "Self-organizing maps," Springer, 1995.
- (38) J. Joutsensalo, "Nonlinear data compression and representation by combining self-organizing map and subspace rule," *ICNN'94*, pp.637-640, 1994.
- (39) Oja E: "Subspace methods for pattern recognition," Research Studies Press, 1983.
- (40) L.Sirovitch and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Am.A*, 2, pp.519-524, 1987.
- (41) M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, 3, pp.71-86, 1991.
- (42) 森崎礼子, 乾敏郎, "顔の認知に関する下側頭葉ニューラルネットワークモデル," 信学技報, NC95-135, 1996.
- (43) Cottrell G.W. and Fleming M.K.: "Face recognition using unsupervised feature extraction," *ICNN'90*, pp.322-325, 1990.
- (44) 鈴木敏, 安藤広志, "2次元射影像から3次元物体の認識と類別モジュール構造を用いた教師なし学習モデル-," 信学論, J79D-II, pp.1291-1300, 1996.

- (45) A.J. O'Toole, H. Abdi, K.A. Deffenbacher and D. Valentine, "Low-dimensional representation of faces in higher dimensions of the face space," *J. Opt. Soc. Am. A*, 10, pp.405-410, 1993.
- (46) D. Valentin, H. Abdi, A.J. O'Toole and G.W. Cottrell, "Connectionist models of face processing: a survey," *Patt. Recog.*, 27, pp.1209-1230, 1994.
- (47) P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *ECCV'96, Lect. Notes Comput. Sci.*, B. Buxton and R. Cipolla (eds.) Springer, pp45-58, 1996.
- (48) D.L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Patt. Anal. mach. Intelli.*, 18, pp.831-836, 1996.
- (49) 池野靖行, 山下幸彦, 小川英光, "相対 KL 変換法によるパターン認識," *信学技報*, PRU95-218, 1996.
- (50) T. Kohonen, G. Nemeth, K-J. Bry, H. Riittinen, "Spectral classification of phonemes by learning subspaces," *ICASSP'79*, pp.97-100, 1978.
- (51) 吉川左紀子, "類似性構造に基づく図形分類反応の検討," *心理学研究*, Vol.51, No.5, pp.267-274, 1980.
- (52) 小松良江, 有木康雄, "部分空間法を用いた向きによらない顔の切り出しと認識," *信学技報*, PRU95-191, 1996.
- (53) Oehler K.L. and Gray R.M.: "Combining image compression and classification using vector quantization", *IEEE Trans. Patt. Anal. Mach. Intelli.*, 17,5,pp.461-473, 1995.
- (54) Pedrycz W.: "Algorithms of fuzzy clustering with partial supervision", *Patt. Recog. Lett.*, 3, pp.13-20, 1985.

- (55) Bensaid A.M., Hall L.O., Bezdek J.C. and Clarke L.P.: "Partially supervised clustering for image segmentation", *Patt. Recog.* 29,5,pp.859-871, 1995.
- (56) Matsunaga H. and Urahama K.: "Partially supervised learning for nearest neighbor classifiers", *IEICE Trans. Inf. & Syst.*, E79-D.2,pp.130-135, 1996.
- (57) 中山弘隆, 谷野哲三: "多目的計画法の理論と応用," 計測自動制御学会, 1994.
- (58) Urahama K. and Furukawa Y.: "Gradient descent learning of nearest neighbor classifiers with outlier rejection", *Patt. Recog.*, 28, pp.761-768, 1995.
- (59) 松永浩之, 浦浜喜一: "クラスタリング回帰による最近傍識別器の教師付き学習法", 信学技報告, PRMU96-80, 1996.
- (60) Duda R.O. and Hart P.E.: "Pattern classification and scene analysis", John Wiley & Sons, 1973.
- (61) Goudail F., Lange E., Iwamoto T., Kyuma K and Otsu N.: "Face recognition system using local autocorrelations and multiscale integration", *IEEE Trans. Pat. Anal. Mach. Intelli.*, 18, 10, pp.1024-1028, 1996.
- (62) 佐藤敦, 山田敬嗣, "一般学習ベクトル量子化の提案," 信学技報, NC95-60, 1995.
- (63) K. Urahama and T.Nagao, "Analog circuit implementation and learning algorithm for nearest neighbor classifiers," *Patt. Recog. Lett.*, 15, pp.723-730, 1994.
- (64) 岡田正之, 浦浜喜一, "クラスタ分割関数回帰の勾配学習法," 信学論, J79-D-II, pp.1913-p1919, 1996.
- (65) 松井, 山下, 若原, 吉室: 文字認識アルゴリズムの複合化手法の検討, 信学技報, PRU92-33, 1992.
- (66) 乾, 山下: 中間視覚における情報統合のメカニズム, 信学技報, NC93-13, 1993.

- (67) 宇野洋二: 視覚・体性間隔と運動制御のモデル, 信学誌, 76, pp.1228-1233, 1993.
- (68) P. Demaritines and J. Hérault: “Curvilinear component analysis: a self-organizing neural networks for nonlinear mapping of data sets,” IEEE Trans. Neural Networks, 8, pp.148-154, 1997.
- (69) 山崎, 石川: センサフージョン, コロナ社, 1992.
- (70) H. McGurk and J. MacDonald, :“Hearing lips and seeing voices,” Nature, 264, pp.746-748, 1976.
- (71) D.W.Massaro and D. Friedman: “Models of integration given multiple sources of information,” Psychol. Rev., 97, pp.225-252, 1990.
- (72) J. Robert-Ribes, J.-L. Schwartz and P. Escudier: “A comparison of models for fusion of the auditory and visual sensors in speech perception,” Artif. Intelli. Rev., 9, pp.323-346, 1995.
- (73) 丹野, 二見, 加納, 星宮: “特徴マップを用いた視聴覚統合モデルによる McGurk 効果の検討,” 信学技報, NC96-133, 1997.
- (74) V. de Sa: “Minimizing disagreement for self-supervised classification,” 1993 Connectionist Models Summer School, pp.300-307, 1993.
- (75) 柴田, 岡部: “相関情報抽出ネットと空間認識能力の教師なし学習,” 日本神経回路学会誌, 3, pp.11-16, 1996.
- (76) 山内, 太田, 石井: “異種センサ情報の統合によるカテゴリーの教師なし学習,” 信学技報, NC95-160, 1996.
- (77) 赤穂昭太郎, 長谷川修, 吉村隆, 麻生英樹, 速水悟: “EM法を用いた複数情報源からの概念獲得,” 信学技報, PRMU96-91, 1996.
- (78) 竹内啓(編): “統計学辞典,” 東洋経済新報社, 1989.

- (79) B.Stein and M.A.Meredith: "The merging of senses," MIT Press, 1993.
- (80) 積山, 東倉: "読唇情報が音声知覚に果たす役割," テレビ技報, 13, pp.31-36, 1989.

目 次

1.1	Muller-Lyer 図形	6
2.1	パターン認識システム	14
3.1	データ例.	29
3.2	画像復元の例.	30
3.3	1次元データの復元例.	32
3.4	セグメンテーションの例.	33
3.5	α の減少によるエッジのなまり方.	34
3.6	マルチスケール平滑化.	35
3.7	マルチスケール平滑化とセグメンテーション.	37
3.8	スパースなデータからの復元.	40
3.9	スパースデータのセグメンテーション.	41
3.10	ドットパターンのクラスタリング.	42
4.1	ドットのクラスタリング	48
4.2	ドット結合のヒステリシス	49
4.3	Muller-Lyer 図形	50
4.4	ドットの変位のシミュレーション結果	51
4.5	矢線の角度による錯視量の変化	52
4.6	矢線の長さによる錯視量の変化	52
4.7	逆錯視図形	53
4.8	同心円錯視のシミュレーション結果	57

4.9	$T = 0$ のときの図 10(a)	58
4.10	分割距離錯視図形	58
4.11	ドット数による左右の長さの変化	59
4.12	曲率の錯視	60
4.13	Ebbinghaus 図形	61
4.14	Delboeuf 図形	62
4.15	Fick 図形	63
4.16	Helmholtz 四角形	64
4.17	Lipps 図形	65
4.18	角度方向錯視図形	66
4.19	平行線幅錯視図形	67
4.20	湾曲錯視図形	68
5.1	代表点の配置 1	76
5.2	代表点の配置 2	78
6.1	簡単なデータによる学習結果 (教師なし 1)	86
6.2	顔データによる学習結果 (教師なし 1)	87
6.3	簡単なデータによる学習結果 (教師なし 2)	90
6.4	顔データによる学習結果 (教師なし 2)	91
6.5	簡単なデータによる学習結果 (教師なし 3)	92
6.6	顔データによる学習結果 (教師なし 3)	93
6.7	簡単なデータによる学習結果 (CLAFIC)	95
6.8	顔データによる学習結果 (CLAFIC)	96
6.9	簡単なデータによる学習結果 (教師あり 1)	99
6.10	顔データによる学習結果 (教師あり 1)	99
6.11	簡単なデータによる学習結果 (教師あり 2)	100
6.12	顔データによる学習結果 (教師あり 2)	100
6.13	簡単なデータによる学習結果 (教師あり 3)	101

6.14	顔データによる学習結果 (教師あり 3)	101
7.1	学習データ	109
7.2	量子化誤差 e_1 と識別誤差 e_2 の α による変化	110
7.3	教師なし学習の代表点 (口) と識別境界 (点線)	111
7.4	全データによる教師あり学習	111
7.5	2 個のデータによる教師あり学習	112
7.6	混合学習の結果	112
7.7	顔画像データの教師なし学習	113
7.8	顔画像データの教師あり学習	113
7.9	4 個の顔画像データによる教師あり学習	114
7.10	顔画像データによる混合学習の結果	114
8.1	式 (8.2) のネットワーク表示	121
8.2	実験例 1	121
8.3	実験例 2	122
8.4	実験例 3	124
8.5	図 8.4 での $p(d_i k)$ の値	125
8.6	図 8.5 の 3 つの $p(d_i k)$ の積	126
8.7	図 8.4 のデータをいくつか入れ換えたもの	127
8.8	図 8.7 での $p(d_i k)$ の値	128
8.9	図 8.8 での 3 つの $p(d_i k)$ の積	129
8.10	$\epsilon = 0$ のときの $p(d_i k)$	130
8.11	図 8.10 での 3 つの $p(d_i k)$ の積	131
8.12	マガーク効果の実験例	133
8.13	出力 $\prod_{i=1}^2 p(d_i k)$ の値	133
8.14	視聴覚入力が逆のときの出力の値	134
8.15	脳の MRI 画像	137
8.16	図 8.15 の画像のセグメンテーションの結果	138

8.17 R,G,B それぞれにノイズを加えた house 画像	139
8.18 図 8.17 の画像のセグメンテーションの結果	140

表 目 次

5.1	2次元関数による分割の結果	80
5.2	1次元関数による分割の結果	81