Some Studies on Linear Regression Problems in Sparse Superposition Codes and Neural Networks

武石, 啓成

出版情報：Kyushu University, 2021, 博士（工学）, 課程博士
バージョン：
権利関係：

（様式 2）

氏　　名　：武石 啓成

論 文 名　：Some Studies on Linear Regression Problems in Sparse Superposition Codes and Neural Networks
（スパース重ね合わせ符号とニューラルネットワークにおける線形回帰問題の研究）

区　　分　：甲

論 文 内 容 の 要 旨

線形回帰問題は，与えられたデータの中に潜む構造を調べる最も基本的な統計的手法の一つであり，多くの領域で応用されている．本論文は，誤り訂正符号とニューラルネットワークにおける線形回帰問題の基礎的問題について考察した結果について述べたものである．

誤り訂正符号については，スパース学習を応用した加法白色ガウス(AWGN; Additive White Gaussian Noise)通信路のための誤り訂正符号であるスパース重ね合わせ符号を取り上げる．スパース学習とは, 線形回帰問題において, 仮説として多数の基底の線形結合を用いたモデルを採用し，真のモデルがスパースであるという仮定の下での学習問題やそれを解くための手法のことを指す．Barron らによって提案されたスパース重ね合わせ符号は，辞書と呼ばれる行列(線形回帰問題における計画行列に対応)の要素のスパースな線形結合で構成したベクトルにより通報を表す. このとき, 復号はスパースベクトルを推定することで行えるため，スパース学習の問題とみなせる．この符号は，辞書(計画行列)の各要素を正規分布に従って独立に生成した場合に，Shannon 限界（通信路容量，すなわち通信速度の理論限界）を達成することが証明されている(Joseph and Barron 2012)．ところが実際のデバイスで用いる場合，厳密に正規分布に従って辞書を生成することが出来ないことが課題である．これに対し，辞書の要素を 2 値に限定し，一様な Bernoulli 分布から生成する設定にまで単純化しても Shannon 限界に達成できるであろうとの予想が Joseph and Barron によって提出されていたが，その証明は困難と思われていた．

深層学習は脳の神経回路を模したニューラルネットワークを多層に重ねたモデルを用いるが，最終層の重みパラメータの学習は線形回帰問題そのものである．深層学習は従来手法に比べて飛躍的に学習精度が優れている一方，その理論的な根拠は未解明である．理論的な課題の一つとして，従来の学習理論の枠組みでは説明出来ない「過剰パラメータの問題」があり，様々な切り口から解析がなされている．こうした研究で扱っている過剰パラメータ問題のいくつかは，最終層の重みパラメータの学習に限っても現れる問題であることが知られており，線形回帰問題の考察により重要な知見が得られる可能性がある．

こうした背景の下，線形回帰問題の計画行列の性質を検討することで得た 3 つの成果について述べる．

第一に，スパース重ね合わせ符号の辞書について提出されていた，Bernoulli 分布に従う辞書を用いても Shannon 限界が達成できるという予想を，最尤復号の場合について肯定的に解決した．この予想がなされた理由は，ベルヌーイ分布を多数重ね合わせた分布(二項分布)は中心極限定理により正規分布に近づくためである．しかし Joseph and Barron によって行われたガウス分布の辞書に関する証明に対して辞書の分布をベルヌーイ分布に置き換えて，二項分布に関する通常の中心極

限定理を適用するだけでは，望む結果が得られなかった．そこでこの証明に合わせて中心極限定理に関する新たな命題を導き，従来の証明に適用することにより，予想の証明に成功した．具体的には，符号長を n とするとき，通信路容量と伝送速度の差分が，$(\log n)^{1/2}/n^{1/8}$ 以上あると復号誤り確率がゼロに近づくことを示した．

第二に，先に証明したベルヌーイ分布に従う辞書を用いた場合の復号誤り確率の上界に関する証明を詳細に検討し，上界を改善できることを証明した．第一の結果を得るための鍵は，二項分布の中心極限定理に関する新たな命題であった．しかし，上界の証明の中に，二項分布の分布関数の値をリーマン積分によって近似する過程があり，その近似精度が不十分であることが判明した．そこで著者は，この近似に Euler–Maclaurin の公式を使うことで，より精密なオーダーの上界を示すことに成功した．このとき，よく知られた公式ではなく，特殊な改良を加えた式を用いた．それ加え，他の証明過程も再検討することで，係数も含めて大きく改良した上界を得ることができた．結果として，通信路容量と伝送速度の差分が，$1/n^{1/4}$ 以上であれば復号誤り確率がゼロに近づくことを明らかにした．これらのスパース重ね合わせ符号に関する結果は，いずれも計画行列の要素が独立に Bernoulli 分布に従う点に特徴がある．

第三に，深層学習において最終層の重みパラメータだけを学習対象とする問題に対し，その Fisher 情報行列がもつ興味深い性質を明らかにした．Fisher 情報行列は，汎化誤差を小さくするために，重視すべきパラメータの更新方向を特徴づける指標である．著者は，深層ニューラルネットの理論解析において近年取り上げられている，Rectified Linear Unit (ReLU)を活性化関数とする中間層が 1 層のニューラルネットの Fisher 情報行列の固有空間を解析した．その結果，入力信号の次元が d であるとき，1 つの大きな固有値に対応するベクトルと，d 個の主要な固有ベクトルと，$d(d+1)/2-1$ 個の次に重要な固有ベクトルが存在し，これらは中間層への重み行列の行ベクトルおよびその対の Hadamard 積に対応することを明らかにした．これは一般のニューラルネットの理論解析への大きな手がかりとなると期待される．また，この結果は，この場合の計画行列の確率分布が，中間層への重み行列と ReLU によって特徴づけられることに対応している．

これらすべての結果は，線形回帰問題の計画行列の特徴と関係するものであり，共通した数学的構造に関する新たな知見として位置付けることができる．

----------以下、提出時には削除してください----------

［作成要領］

1. 用紙はＡ４判上質紙を使用すること。

2. 本文の文字サイズは１０．５ポイント（「論文内容の要旨」の文字は１２ポイント）
１行の字数４４字、行数４２行、余白（左右２０㎜・上下２５㎜程度）をあけ、頁数は記入しない。

3. 論文内容の要旨は、A4サイズ1頁に2000字程度（最大２頁以内を目安）にまとめる。

4. 図表・図式等は随意に使用する。

5. 氏名は、外国人の場合，カタカナ表記（漢字圏の学生は漢字）で記入する。
※氏名の順番等は，パスポートや外国人登録証と照合，学籍氏名も含めてできるだけ統一させる。

6. 主論文等の題名と合わせる。外国語の場合，字体・文字の大小について統一させ，和訳を（ ）
カッコ書きで付記する。

7. 甲・乙を明示すること。

※ この原稿は、電子データも提出する。