—— JARIP 学会賞受賞記念論文 ————————————

深層学習の近似能力について

楠岡成雄 *

2020 年 9 月 4 日投稿

1　はじめに

機械学習には、対象や目的に応じて多くの手法が存在する。その代表的なものとして、教師あり学習、教師なし学習、強化学習がある。どの場合も関数近似、則ち未知の関数を何らかの関数で近似するということが必要となる。

機械学習を厳密に定義することは大変難しいが、大雑把に言って、集合 $\tilde{\mathcal{X}}$ を定義域とし、集合 $\tilde{\mathcal{Y}}$ を値域とする写像 $\tilde{f}:\tilde{\mathcal{X}}\to\tilde{\mathcal{Y}}$ に対して、それを近似する写像を機械的に見つける手法を指す。ただし、定義域 $\tilde{\mathcal{X}}$ 、値域 $\tilde{\mathcal{Y}}$ は未知であることが多く、集合 $\tilde{\mathcal{X}}$ を含む既知である集合 $\mathcal{X}$、集合 $\tilde{\mathcal{Y}}$ を含む既知である集合 $\mathcal{Y}$ を考え、さらに写像 $\tilde{f}:\tilde{\mathcal{X}}\to\tilde{\mathcal{Y}}$ の拡張である写像 $f:\mathcal{X}\to\mathcal{Y}$ を考え、写像 f を近似する写像を見つけるということが一般に行われる。写像 f を近似する方法として、多くの場合用いられるのが、パラメータ θ 付きの写像 $F(\cdot,\theta):\mathcal{X}\to\mathcal{Y}$ を考え、パラメータ θ を動かして、写像 f をよく近似する写像を見つけるという方法である。どのようなパラメータ付き写像のクラスを考えるかについても様々な方法が提案されているが、本論文では、深層学習と呼ばれる写像のクラスについて考えていく。特に、パラメータ数を増やすことで、どのような写像がどのような意味で、その写像のクラスで近似できるかという問題について考察する。

本論文では一般論を扱うが、例がないとわかりにくいと思うので、教師あり学習を例として考えていく。教師あり学習においては、先ず既知データとして、入力・出力のデータの組の列 $(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ が与えられている。簡単のために入力 $x_k, k=1,\ldots,n,$ は N 次元ベクトル、出力 $y_k, k=1,\ldots,n,$ は M 次元ベクトルとする。

考える例として、以下の 2 つの例を考える。

例 1 (犬猫判定問題)

これは、jpeg で記述された画像データを見て、それが犬か猫かを判定するという問題である。$\tilde{\mathcal{X}}$ は犬またはネコの jpeg データの全体となるが、これを確定させせることは困難である。犬、猫に限らず考え得る可能なすべての jpeg データは $[0,1]^N$ に埋め込むことができるので $\mathcal{X}=[0,1]^N$ とすればよい（白黒の画像データであれば N は素子数とすればよい)。$\tilde{\mathcal{Y}}$ は今の場合 { 犬、猫 } であるので、犬に対して 0, 猫に対して 1 を割り当てることにすれば、$\tilde{\mathcal{Y}}=\{0,1\}$ とおけばよく、$M=1$ である。$\tilde{f}$ は $x\in\tilde{\mathcal{X}}$ が犬のデータであれば $\tilde{f}(x)=0$, 猫のデータであれば $\tilde{f}(x)=1$ という写像である。写像 f は写像 $\tilde{f}$ の拡張であれば何でもよい。

例 2 (時系列データに基づく予測問題)

$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ が時系列データであり、将来の出力データを予測するという問題を考える。各時点 m での入力に対応する N 次元のベクトル値確率変数を X_m 出力に対応する M 次元のベクトル値確率変数を Y_m とし、$N+M$ 次元のベクトル値の確率過程 $\{(X_m,Y_m);\ m\in\mathbf{Z}\}$ が定常過程であり、$(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)$ が $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ の実現値と考えられる場合を考察する。

過去の入力データ $X_k,\ k=1,\ldots,n,$ から出力 Y_{n+1} を推定しようとすると、二乗損失関数の下では、

$$E[Y_{n+1}|X_n,.X_{n-1},,\ldots,X_1]=\tilde{f}_n(\{X_{n+k}\}_{k=0,-1,\ldots,-n+1})$$

*明治大学客員教授、東京大学名誉教授