

機械学習を用いた順位予測モデルの構築と精度評価
-競馬予測を事例として-岩崎 照英(21X4009) 田中 幸佑(21X4038)
徳丸 颯太(21X4102) 指導教員 劉 慶豊

1. 研究の背景と目的

競馬は多くの要因が絡み合う複雑な競技であり、従来の予測手法では多くの不確実性が残されている。一方では、近年のデジタル化と AI 技術の進展により、膨大なデータを用いた分析が可能となり、より精度の高い予測が期待されている[1]。本研究では、まず競馬の基礎知識を解説し、現在進行している AI による競馬予測の動向を紹介する。その後、過去のレースデータを利用して、機械学習の手法ニューラルネットワークで予測を行った。さらに、その予測結果を基に、ベット戦略を実行して投資収益率(ROD)の向上を実現した。

2. 利用する機械学習の手法

本研究では機械学習の一種であるニューラルネットワークを利用して分析する。モデルは入力層、隠れ層、出力層の3種類の層から構成される。各層には「ノード」と呼ばれる計算単位が含まれており、これらのノードが互いに繋がりが合うことでネットワークを通じてデータの流れが生まれる構造になっている。それぞれのノードは数値を受け取り、その数値に対して重みやバイアスを適用し、さらに活性化関数を通して処理を行う。

・入力層

モデルの一番外側に位置し、学習対象のデータがまずここに入力される。画像データの場合はピクセル情報、テキストデータの場合は数値ベクトルなどが割り当てられる。

・隠れ層

入力層から送られてきたデータを受け取り、特徴を学習する層である。通常、ニューラルネットワークの性能は隠れ層の構成に大きく依存する。例えば、隠れ層の数が多いほどモデルはデータのより複雑なパターンを学習できるが、過剰に層を増やすと計算コストが上がり、過学習のリスクも高まる。

・出力層

最終的な予測結果を出力する層である。分類タスクではソフトマックス関数などを使い、各クラスの確率を出力することが多い。回帰タス

クの場合は数値がそのまま予測値となって出力される。

・活性化関数

各ノードの出力を非線形にするために使用され、ネットワークの学習を可能にする重要な要素である。

3. モデル設定

Keras の Sequential API を使用して表 1 のニューラルネットワークモデルを構築した。

表 1：モデルの構成

層	ユニット数	関数
入力層	256 層	Dense 層に ReLU 活性化関数
隠れ層	128 層	Dense 層、過学習対策として Dropout 層、ドロップアウト率 30%
出力層	1 層	Sigmoid 活性化関数

4. 分析とその結果

4.1 実装

訓練データを用い、エポック数 100 でモデルを訓練した。図 1 で訓練の進行中に、検証データで精度と損失の推移を観察した。ニューラルネットワークの学習では損失関数の値を小さくするように学習が進んでいくため、一般には損失値が小さいほどいいモデルであると考えられる。図 1 は 100 エポック目までバリデーションデータの損失値の傾向が継続的に下がっているため、過学習は発生していないといえる。

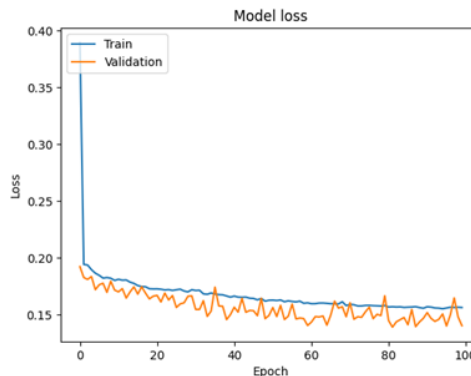


図 1：訓練データの精度の推移

各特徴量が競馬予測において重要であるかどうかについて評価するために、Permutation

Importance (PI)を計算した。PI とは、機械学習モデルにおいて、特定の特徴量が予測のためにどの程度重要であるかを評価する指標である。各特徴量の重要度を図 2 に示した。

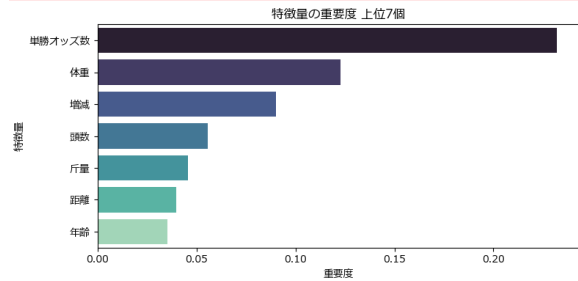


図 2：特徴量上位 7 個

図 2 の分析の結果によれば、「単勝オッズ数」の重要度は約0.23という値になり、他の項目と大きな差をつけた。競馬において「単勝オッズ数」はそのまま競走馬の人気順となり、オッズが低ければ低いほど 1 着になる可能性が高くなるということから「単勝オッズ数」の重要度が理解できる。また、2, 3 位は「体重」、「増減」となっており馬の体重に関する項目が上位を独占する形となった。

4.2 ベット戦略一

ニューラルネットワークの予測結果に基づき、全てのレースで 1 着になると予測した馬に同じ金額を賭け、実際のオッズと組み合わせで式(1)で ROI を算出した。

結果として、総ベット額375,900円に対して、総払い戻し額409,730円となり、ROI は0.09となった。

$$ROI = \frac{\text{総払い戻し額} - \text{総ベット額}}{\text{総ベット額}} \quad (1).$$

4.3 ベット戦略二

ベット戦略二では戦略一と同じニューラルネットワークのモデルの予測をもとに、各ベットの期待値を計算し、期待値が高いベットに資金を配分し利益を最大化するベット戦略を採用した。この戦略により、ROI は0.27となり、ベット戦略一の0.09を大幅に上回る結果を達成した。この ROI は日本の競馬の ROI の平均と言われている ROI は-0.136[2]に比べて、高い数値となった。このことより、機械学習による予測は ROI の改善に大きく貢献していると言える。

5. 結論

本研究では、AI 技術を活用した競馬予測における予測精度の向上を目指し、ニューラルネ

ットワークを用いたアプローチを検討した。特に、単勝式において、ニューラルネットワークの予測結果に基づいた戦略によって ROI の改善が確認されたことで、ニューラルネットワークが競馬予測における重要なツールとなり得ることが示唆された。

一方では、ニューラルネットワークを用いた競馬予測には計算コストの課題が残る。トレーニングには大量のデータと計算リソースを必要とするため、計算負荷が高い。この課題を解決すべく、今後の研究においては、ニューラルネットワークをさらに発展させることが重要である。また、強化学習 (Reinforcement Learning) は、環境との相互作用を通じて予測を改善していくアプローチであり、競馬予測における動的な要因を学習するのに適していると考えられる。そのため、別のアプローチとして、ニューラルネットワークと強化学習との統合が有望であろう。

参考文献

- [1] 杉山 航太, 山下 遥, 部分再帰型ニューラルネットワークを用いた最適な競馬の betting 戦略, (2023).
- [2] 石黒秀明. (2021). 偶然性を排除した馬券購入の「独自のノウハウ」の存在について: 馬券回収率に対する統計的仮説検定による実証研究 [東京高裁令和 2.11. 4 判決]. 税理: 税理士と関与先のための総合誌/ぎょうせい編; 日本税理士会連合会監修, 64(3), 210-222.