

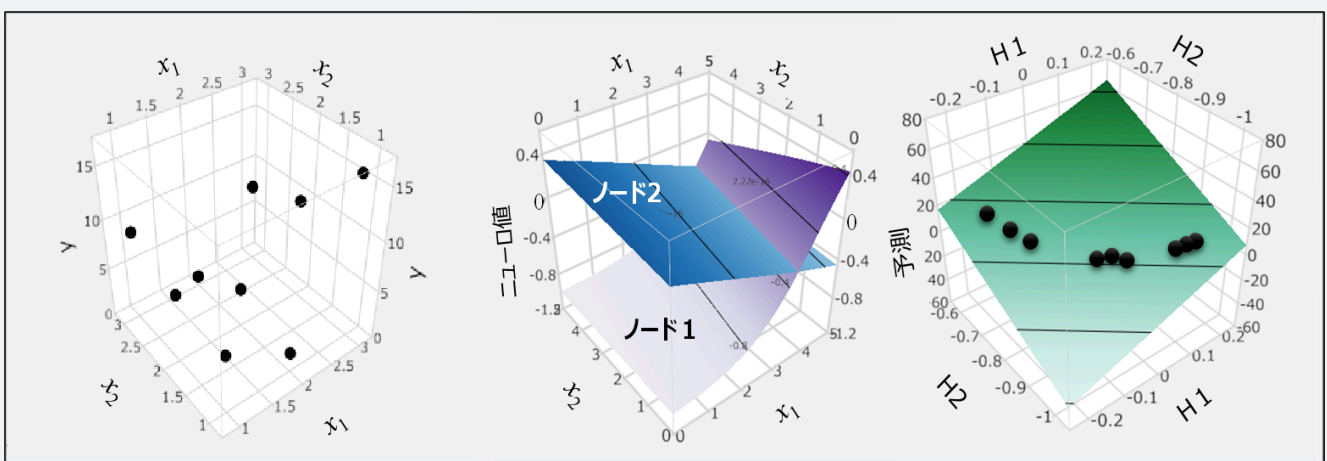
連載 統計的機械学習ことはじめ

廣野 元久 著

第5回 ニューラルネットワークによる量的特性の予測

活性化関数にロジスティック関数を用いるニューロ判別では、質的な特性の予測値のデータ領域は0～1に収まるように変換され、通常は0.5を境界値として2つのクラスに分類されます。量を予測する特性では少し工夫が必要です。まず、活性化関数を使って予測値の最大値を1に、最小値を0あるいは-1になるように計算します。ついで、得られた値を元の特性のデータ範囲に戻します。今回は量的な特性を予測するニューラルネットワークを重回帰モデルと比較しながら説明します。活性化関数は色々提案されていますが、線形式と非線形なTanH関数 $h(a)=(\exp(2a)-1)/(\exp(2a)+1)$, $-1 \leq h(a) \leq 1$ を使います。

最初に、隠れ層が1つでノード数が2の場合のニューラルネットワークを考えます。説明は後回しにしますのでグラフでイメージをつかんでください。図10左はあるデータの三次元散布図です。中央の図はニューロ値 $h_f(a)$ を x_1 と x_2 で計算したものです。縦軸のスケール ($-1 \leq h_f(a) \leq 1$) に注意してください。この散布図には x_1 と x_2 を入力して計算された2つのノードの曲面が表示されています。曲面上に予測された等高線が実線で表示されています。右の図は各ノード (H1とH2) を説明変数として y を重回帰モデルで予測した平面と特性の観測値 (●の点) を表示しています。また、平面上に予測された等高線が実線で表示されています。予測値は元のデータ範囲に戻っていることにも注意してください。



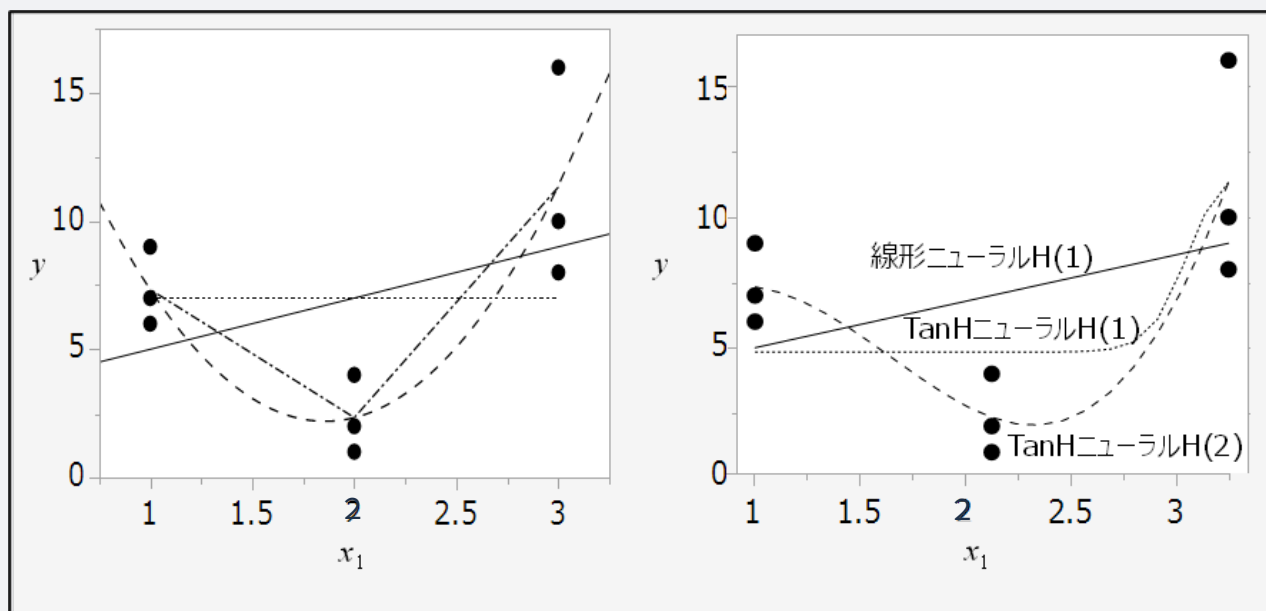
x_1 と x_2 , y の散布図

x_1 と x_2 、ニューロ値の散布図

H1 と H2、予測値の散布図

図10 あるデータセットの三次元散布図

以下ではどのようにして入力 (x_1 と x_2) から隠れ層を通じて予測値を算出しているのかを説明します。まず、 x_1 で y を予測することを考えます。図11左は x_1 と y の散布図に点線で平均線を、実線で1次式を、破線で2次式を、一点鎖線で x_1 の水準平均を結んだ折れ線をあてはめています。線形なモデルでは2次式、あるいは x_1 の水準平均をあてはめるのがよいことがわかります。一方、右の図は x_1 と y の散布図に直線で隠れ層1ノード1の線形ニューラルH(1)モデル、点線で隠れ層1ノード1の TanH 関数を用いた TanH ニューラルH(1)モデル、破線で隠れ層1ノード2の TanH 関数を用いた TanH ニューラルH(2)モデルをあてはめたものです。TanH 関数を2つにすることにより、下に凸の曲線のあてはめが可能になります。左右の破線の曲線を見比べると、線形2次式は $x_1=2$ で左右対称となります。一方、2つの TanH 関数を使ったニューラルネットワークのモデルは、非対称のモデルを作れるため推定精度が向上します。



線形モデルのあてはめ

ニューラルネットワークのあてはめ

図11 線形モデルとニューラルネットワークの比較

では、ニューラルネットワークはどのような計算を行っているのでしょうか。図12は $x_1=1$ を入力した場合の隠れ層1ノード1のニューラルネットワークの計算例です。重み w と閾値 θ の推定は複雑なので計算機を使いますので説明を省きます。図の左が線形ニューラルの計算結果です。複雑な計算をしていますが、結局は単回帰モデルと同じになります。つまり、 $-1.853-9.181(-0.529-0.218x_1)$ を展開すれば、単回帰モデル $3+2x_1$ になるからです。右の図は TanH ニューラルの計算例です。 $-18.393+6.527x_1$ から $x_1=1$ を入力すれば、 -11.866 となるので、この値を TanH 関数に入れると -1 と求まります。これより、

$8.385 + 3.552x_1$ の x_1 に -1 を代入して 4.833 が得られます。2度、単回帰分析を繰り返していることがわかります。

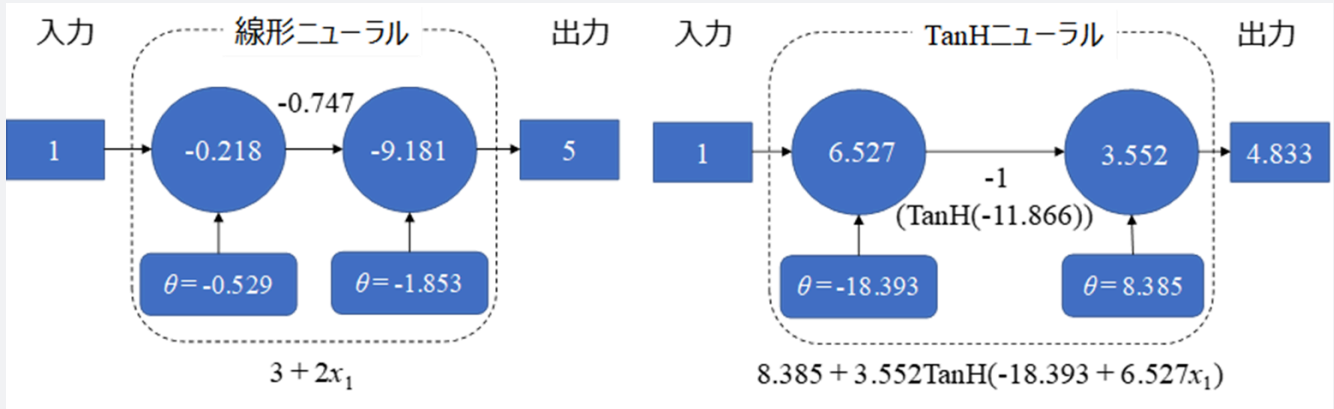
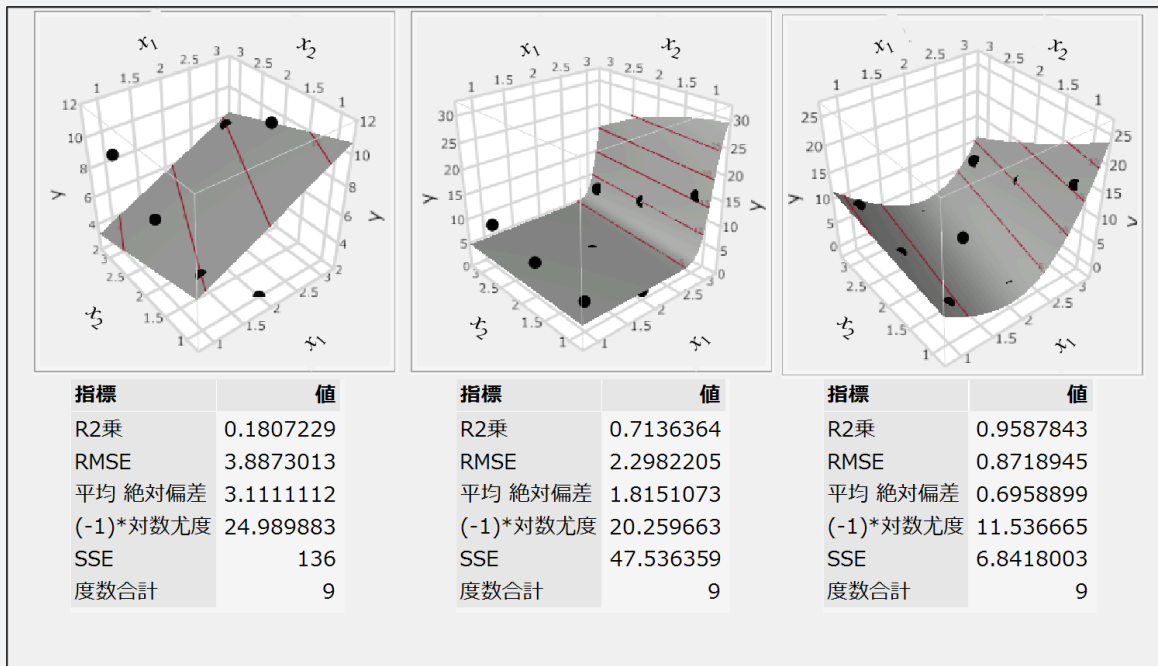


図12 入力1 隠れ層1ノード1のニューラルネットワークの計算例

ノード数が1の2つのモデルではどちらもあてはまりはよくありません。そこで、入力に x_2 を追加します。図13左は隠れ層1ノード1の線形ニューラルモデルの結果です。これは主効果のみの線形重回帰モデルになっています。推定された平面に引かれた線が y の予測値の等高線です。線形なモデルの寄与率 R^2 は0.18と小さい値です。中央の図は隠れ層1ノード1の TanH ニューラルモデルの結果です。推定された曲面から、 x_1 の値が3に近づくにつれて急激に y の予測値が大きくなっていることが読み取れます。 x_2 が y に与える影響は小さいのですが、寄与率 R^2 は0.71と改善されています。TanH ニューラルでノード数を1つ増やすと推定精度がどう向上するかを右の図に示します。下に凸の曲面が得られています。主効果と2次項を入れた重回帰モデルとよく似た結果になっています。寄与率は主効果と2次項を入れた重回帰モデルよりも僅かに優勢で $R^2=0.96$ と向上しています。

以上から線形ニューラルはノードごとに平面を形成し、主効果のみの重回帰モデルであることがわかります。TanH ニューラルに代表される非線形ニューラルはノードごとに曲面を形成するモデルを経由して、得られたニューロ値 H_i で特性を予測する重回帰モデルになっています。言い換えると、各ノードで形成された複数の非線形回帰モデルの重み付き和を使って、出力のデコボコな値に滑らかな曲線で予測を行うモデルということです。



線形ニューラル ノード1のTanHニューラル ノード2のTanHニューラル

図13 x_1 と x_2 を使った隠れ層1の三次元散布図と統計量

最後に、図10に戻ると、中央の図は x_1 と x_2 を使って、各ノード H1 と H2 の曲面を示したものです。ノード1は x_1 の値が大きくなるにつれて指数関数的に大きくなる曲面、 $H1 = \text{TanH}(0.242 - 0.146x_1 + 0.025x_2)$ 、ノード2は x_1 の値が大きくなるにつれて右下がりの直線に近い曲面、 $H2 = \text{TanH}(-4.275 + 0.548x_1 - 0.144x_2)$ が描かれていたのです。右の図は H1 と H2 を使って y を予測する平面、 $\hat{y} = 104.565 + 96.269H1 + 115.932H2$ が描かれていました。重みの値が大きいのはニューロンの世界を観測値の世界に引き延ばすためです。簡単に申し上げますと、特性 y について $y^* = (y - y_{min}) / (y_{max} - y_{min})$ の変換を行い、 y^* を予測する重回帰モデルです。ニューラルネットワークは出力をもっとも上手く予測する曲面を作るために、複数のノードの曲面の重み付け和のパラメータを推定する方法です。隠れ層が複数の場合も同様で、曲面の重み付き和を幾重にも重ねた曲面で特性を予測しています。注意すべき点は、ニューラルネットワークは予測の方法で因果の説明や解釈には不向きということです。隠れ層の階層やノードが増えるにつれて複雑な情報の伝播モデルとなり、解釈ができなくなるからです。また、ニューラルネットワークでは初期値の与え方によりパラメータの推定値が変わるので、何度か繰り返し計算を行い、モデルの安定性を調べる必要があります。



著者紹介

廣野 元久 (ひろの もとひさ)

1984年(株)リコー入社。以来、社内の品質マネジメント・信頼性管理の業務、SQCの啓蒙普及に従事、品質本部QM推進室長、NA事業部SF事業センター所長を経て、現在、(株)リコー倫理審査委員会委員。

東京理科大学工学部経営工学科 非常勤講師 (1997~1998年)、慶應義塾大学総合政策学部 非常勤講師 (2000~2004年)。(一財)日本科学技術連盟 多変量解析法運営委員会委員、講師。