

連載 統計的機械学習ことはじめ

廣野 元久 著

第4回 ロジスティック判別を超えてニューロ判別のススメ

多変量解析では、特性が量的変数の予測には、古くから重回帰分析が使われています。特性が質的変数の場合の予測は発生確率を求めることになりますから、ロジスティック回帰分析が使われます。その応用としてロジスティック判別分析が利用できます。ロジスティック判別分析では予測確率の閾値（通常は0.5）によって良・不良といった結果に対する判断を行います。機械学習のニューラルネットワークの立場から眺めると、ロジスティック判別分析は隠れ層が1つでノード（潜在変数）が1つの単純なニューロ判別モデルに相当します。今回はロジスティック判別モデルの拡張がニューロ判別モデルに繋がることを説明します。なお、隠れ層が多層になった複雑なモデルは深層学習に使われます。

ニューロ判別は人間の脳内のニューロンを模倣しています。ニューロンは電気信号として情報伝達を行いますが、シナプスの結合強度によって情報の伝わりやすさを変えています。このモデルは結合強度を重み w_i で表し、入力と出力を結合したものになります。入力と出力を繋ぐ層は隠れ層とよばれ、入力と出力だけでは線形分類しかできない欠点を補います。隠れ層を持つことで柔軟な非線形分類ができるのです。ニューロ判別は隠れ層に配置されたノード内で処理を行います。このとき、入力に対して重み w_i (偏回帰係数に相当)とバイアス b (切片に相当)を使って線形式を作ります。計算結果はそのまま出力されるのではなく、活性化関数(変数変換に相当)を使って計算された値を出力します。活性化関数は様々なものが提案されています。たとえば、ステップ関数： $h(a)=\text{if}(a>0 \Rightarrow 1, \text{else } 0)$ やロジスティック関数： $h(a)=1/[1+\exp(-a)]$ などがあります。ロジスティック関数の場合は、どんな値でも0~1の値で出力されますから、入力の値の桁数が大きくなってもそれほど結果には影響を与えず、誤差が生まれにくいという効果が期待できます。ロジスティック関数はS字型の曲線で、入力が0の場合はちょうど出力が0.5になります。このとき、図7に示すように隠れ層が1つでノードが1つの最も単純なニューロ判別モデルはロジスティック判別モデルに相当します。 $h(a)$ によって予測確率が計算され、その確率が0.5未満であればクラス1に、0.5より大きければクラス2に振り分けられるルールと考えられるからです。

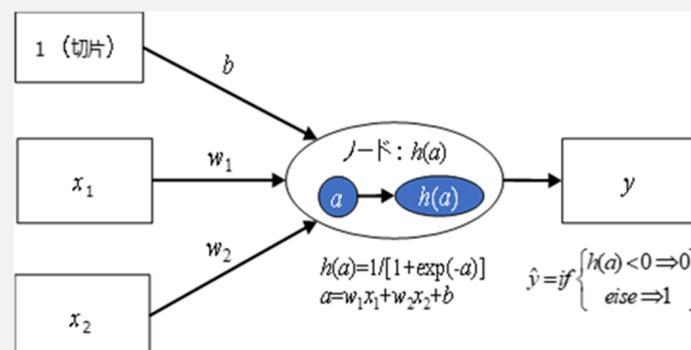


図7 ロジスティック判別モデルをニューロ判別モデルで表した例

では、なぜ隠れ層やノードの数を増やすメリットがあるのでしょうか。図8を使って説明します。図8の左の散布図では X_1 と X_2 の平面に良品○と不良品●がマーカーで区別されています。散布図の領域Aに不良品が集まっています。散布図上の実線がロジスティック判別の境界です。その際の判別結果を左の散布図の下にLR学習という名前で表にしています。誤判別数は200個中の12です。ロジスティック判別の方は手抜きをして全個体での判別結果を示しています。一方、散布図上の破線はニューロ判別の境界線です。この境界は隠れ層1でノード1のニューロ判別モデルの結果です。活性化関数の形が若干、ロジスティック関数とは異なりますが、ほぼ同等の境界が得られています。誤判別は散布図下のNN学習データ（上表）132個中の8で、NN検証データ（下表）では68個中の6となっています。機械学習では学習用データでモデルを作り、検証用データで汎化性を確認します。このため、ニューロ判別では毎回、学習データはランダムに選ばれ、初期値も変化しますので、分析ごとに境界は変化します。もっとよいニューロ判別モデルを探せますが、推定精度の向上は今回の目的ではありません。ここでの目的は、一番単純なニューロ判別のモデル（隠れ層1、ノード1）はロジスティック判別モデルと同等なものであるということを示すことです。また、今回はモデル選択を行いませんから評価用データの設定はありません。

つぎに、図8の真ん中の散布図を見てみましょう。今度はBの領域にも不良が出ています。ロジスティック判別の境界は実線で表していますが、領域Aの不良を判別する能力がありません。領域Bの判別も上手くいったとはいえません。誤判別結果も芳しくありません。不良品の領域が複数ある場合は、隠れ層1、ノード数1の単純なモデルでは限界があるようです。不良の領域AとBに対して別々のロジスティック判別分析を行う必要があったのです。一方、ニューロ判別では少しだけ格上げして隠れ層1のノード数2のモデルを使います。その結果、2本の破線の境界が得られ上手く判別でき、誤判別はNN学習データ（上表）では132個中2、NN検証データ（下表）では68個中4です。

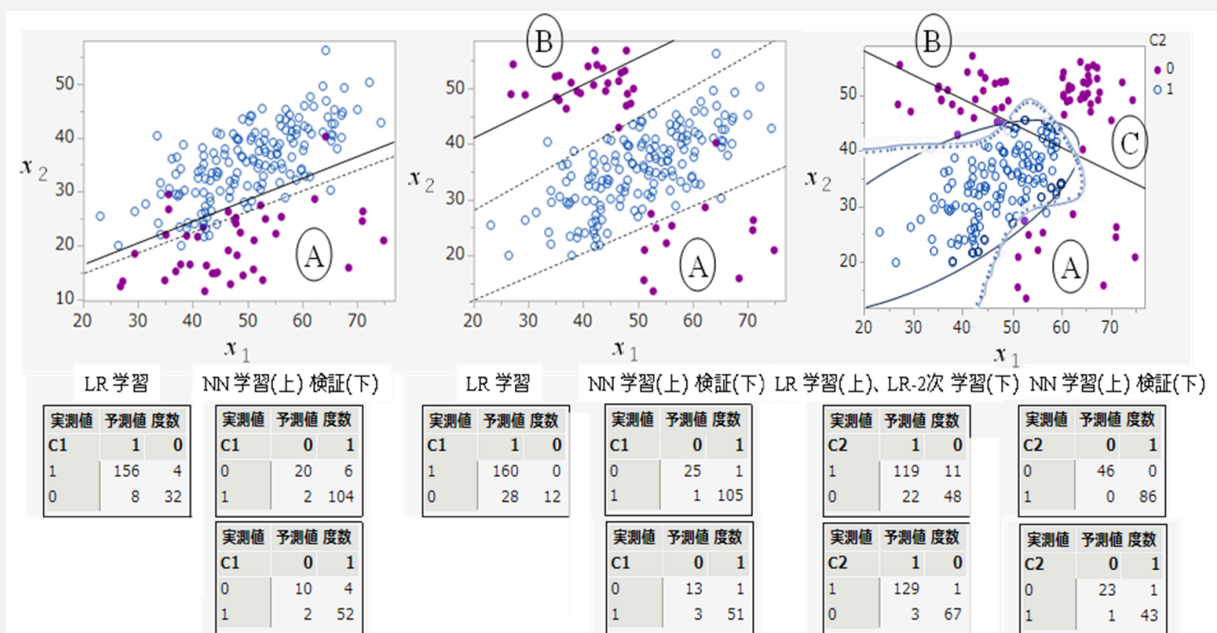


図8 3つの不良領域のパターンでのロジスティック判別とニューロ判別の比較

最後に図8の右側の散布図を見てみましょう。不良はCの領域にも出ています。ロジスティック判別の境界を実線で表しています。Cの領域ではうまく不良品を判別できていますが、Bの領域では判別が甘く、Aの領域では全く判別能力がありません。誤判別はLR学習（上表）から、200個中33もあります。ところで、ロジスティック判別の応用として説明変数側を多項式（ここでは2次項と交互作用項を追加）にすると、図中の楕円の曲線の判別境界が得られます。その際の誤判別はLR-2次学習（下表）のように200個中4と改善されます。図は省略しますが、3次項まで考えるとモデルは不安定ながら誤判別はほぼない複雑な境界が得られます。一方、ニューロ判別の隠れ層1でノード数3のモデルは図中の楕円と似た曲線的な判別境界が得られます。今回は隠れ層2で入力に近い隠れ層のノードを3、出力に近い隠れ層のノードを2にしたモデル（図9）を使います。その結果、破線の複雑な曲線が境界として得られます。誤判別はNN学習データ（上表）では132個中0、NN検証データ（下表）では68個中2となっています。

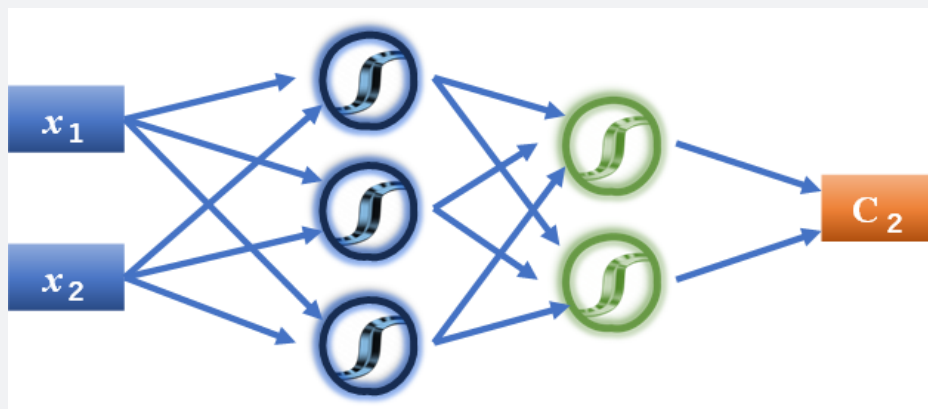


図9 N2N3ダイアグラム（煩雑のため切片項を省略）

ロジスティック判別の応用では説明変数側で多項式を導入することで誤判別を減らすことができますが、説明変数の数が多いとモデル作りは非現実になります。しかし、ニューロ判別では隠れ層とノード数を設定すればモデルの作成は容易で、柔軟な判別境界が得られます（判別境界の解釈容易性は別として）。以上からニューロ判別をロジスティック判別が拡張されたもの、つまり複数の判別境界を作り、それらを繋いだと考えれば、ニューロ判別の敷居が低くなるのではないかと思います。



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。