連載　統計的機械学習ことはじめ

廣野　元久　著

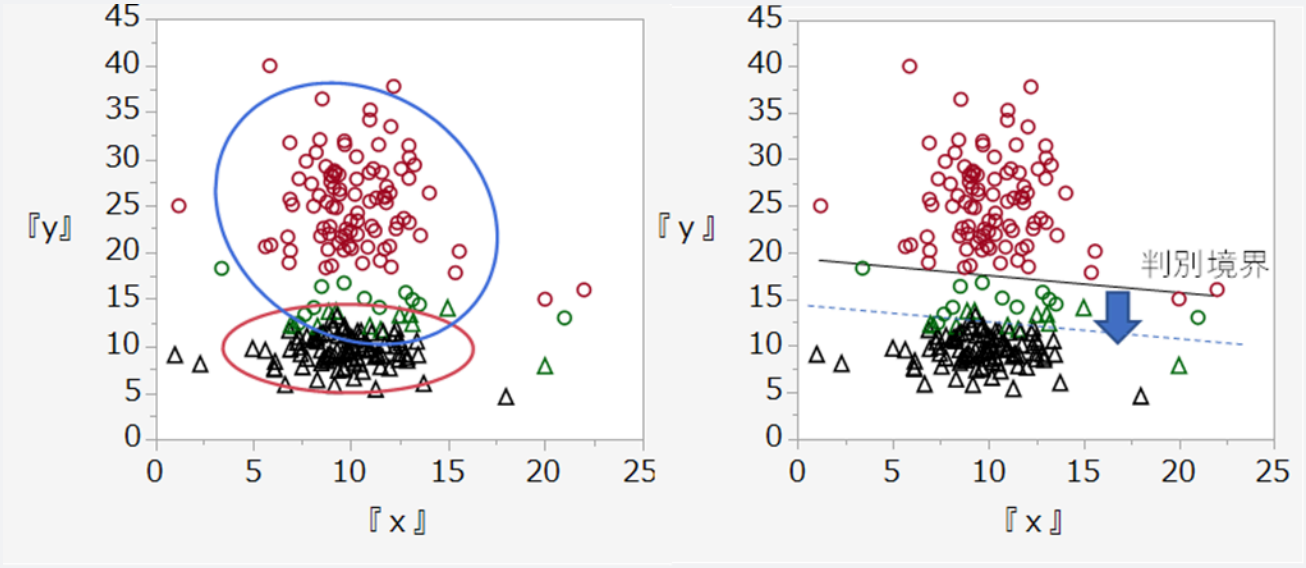
第３回　SVM（サポートベクターマシン）の考え方

判別の目的はグループの境界を決めることですから、グループの境界付近にある点を重視し、境界から遠い点を無視するほうが都合のよい場合があります。その要求に答えてくれるのがSVM（サポートベクターマシン）です。SVMは機械学習の主要なアルゴリズムの１つで、線形な方法とカーネル法を用いる非線形な方法があります。

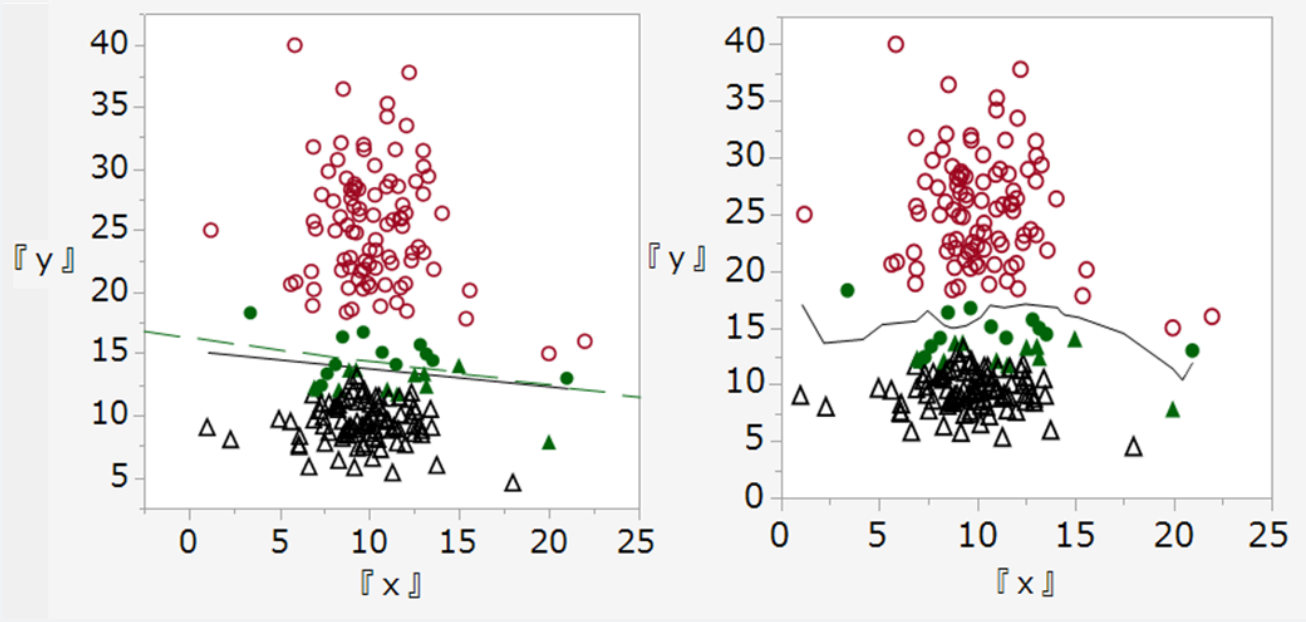
判別分析では、すべての個体を使って判別境界を求めています。観測値は多変量正規分布に従う確率変数であること、また、各グループの母分散と母共分散は等しいという前提に基づく統計モデルです。少しお行儀を崩した線形SVMは、判別に役立つ個体を数理的に選んで、少数の観測点（サポートベクター）で境界を定める方法です。

判別分析が行われる状況で、統計的な前提が崩れた場合にはどのようなことが起きるのかを確認してみましょう。図５-①は層別因子でグループ分けした確率楕円を加えた層別散布図です。層別因子の水準であるｃ１の個体の位置を△で、水準ｃ２の個体の位置を○で分けています。２つの確率楕円の様子から、「２つのクラスに等分散共分散性が成り立っている」と感じられないですよね。統計仮説が崩れている状況で判別境界を求めると何が起こるでしょうか。判別分析で得られた判別境界を追記したグラフを図５-②に示します。ｃ１をｃ２と誤分類することはないのですが、ｃ２をｃ１と誤分類することが多い結果となりました。グラフから直感的に、「判別境界線を垂直方向にもっと下げたほうがよい」と感じるでしょう。そこで、主観的に境界に近い観測点だけを使って判別分析を行い、新たな境界線を引くことを考えます。図５-③の実線は●と▲以外の観測点の重みを０にした場合の判別境界です。すべての観測点を使った判別境界線よりもよい判別結果が得られています。このような恣意的な方法は分析者の主観が色濃く反映されるので、客観性に乏しく好ましい方法とはいえません。図中の破線は線形SVMにより求めた判別境界線です。こちらは客観的な基準で選定された観測点を使って得られたものになります。また、図５-④の複雑な境界は非線形SVMにより求めた判別境界線です。非線形SVMのほうがよりよい判別ができているように感じます。

（次ページへつづく）



①クラスで層別した確率楕円　　②２群の判別境界の追記



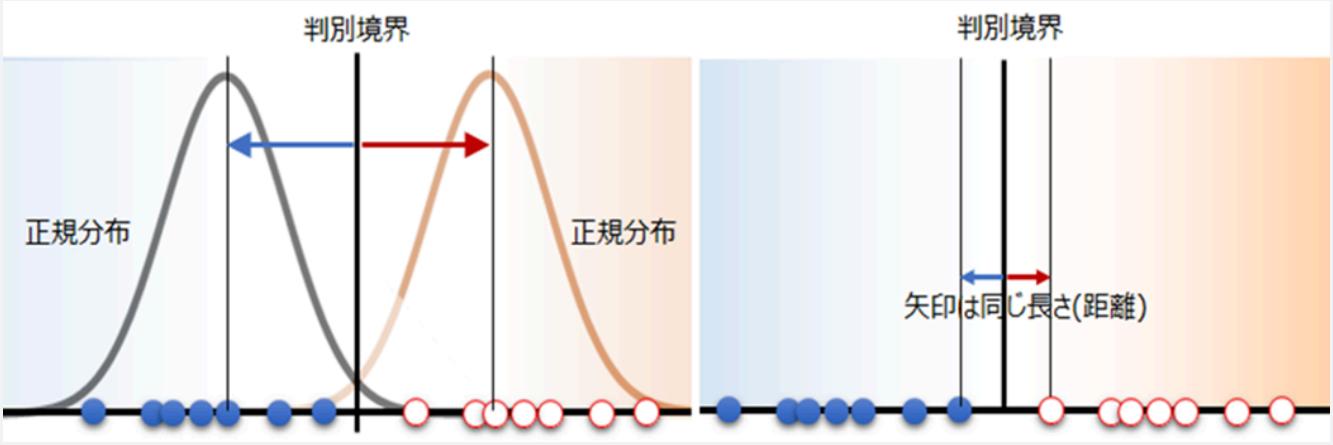
③恣意的な境界と線形SVMの境界　　④非線形SVMの境界

図５　線形判別分析とは異なる方法で求めた判別境界(△はｃ１，○はｃ２)

SVMは広く使われている学習アルゴリズムの１つです。SVMでは分類に必要な個体の情報（分類をサポートするベクトルとして）を使って判別境界を計算します。カーネル法と組み合わせることで、非線形な判別を行うことができ、高い判別性能が得られる方法です。判別分析もSVMも判別関数の符号で、どちらのクラスに所属するかを決定しています。判別分析では図６-①のように正規分布を仮定しており、マハラノビス距離を基準に分類します。判別境界は各クラスの平均からのマハラノビス距離が等しい点になります。SVMでは図６-②のように判別境界は各クラスの1番近い個体である２点からできるだけ距

（次ページへつづく）

離を取るように作られます。これをマージン最大化といいます。この判別境界線に１番近い観測点をサポートベクターといいます。名前の由来は、観測点が境界を支える（サポートしている）ベクトルであることから来ています。一般に、サポートベクターの数が少ないと小さな次元で分類でき、判別境界は単純です。一方、サポートベクターの数が多いと誤分類の数が見た目で少なくなるのですが、高い次元で分類されているので判別境界が複雑になります。サポートベクターの数が必要以上に多い場合は、モデルが過学習している可能性が高く、汎化能力が劣っている可能性があります。



①判別分析による判別境界　　②SVMによる判別境界

図６　2つの判別境界の考え方の比較

ところで、SVMにも弱点があります。カーネル法によって特徴量の数pへの依存は小さくなるのですが、個体数nが多くなるに従い、計算に掛かる時間が膨大になってしまいます。解釈上にも課題があります。非線形なSVMはカーネル法を使った高次元での線形化モデルです。実際の観測点の空間に戻すと複雑な境界を作ることになります。このため、境界は人が理解できる式の形で求めることができなくなります。つまり、良い分類はできるのですが、分類した理由が分析者にもわからないという状況が発生します。このことから、SVMは予測向きで、因果推論や応答の制御には不向きな理由となっています。

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。