

カーネル回帰による判別分析における最適な正則化パラメータの推定

堀美乃里 (指導教員：吉田裕亮)

1 はじめに

統計学上のデータ解析手法のひとつに、判別分析がある。いくつかのグループに分かれているデータを元に、それらが「どういう基準で分けられているのか」という関係を解析することで、分類されていないサンプルがどのグループに属するかを予測する手法である。

判別器のひとつとしてカーネル回帰による判別方法が挙げられる。そこで本研究では、ガウスクERNELを用い、カーネルパラメータ β が与えられたもとの、カーネル回帰における最適な正則化パラメータ λ を推測するひとつの手法を提案する。

2 カーネル回帰

2.1 カーネル法

カーネル法とは、カーネル関数を利用し、観測データを高次元のベクトル空間に写像し、変換後のデータに線形の手法を用いることで、非線形の解析を実現することができるものである。すなわち、カーネル法を用いると複雑な関数を表現でき、非線形な関係を考慮できる。

2.2 カーネル関数

カーネル関数は、特徴量で見たときの x と x' の類似度 (直感的には x と x' の近さ) を表していると考えられることもでき、2つの要素 x, x' に対し、それぞれの特徴ベクトル $\phi(x), \phi(x')$ の内積として定義される。すなわち、 $\phi(x), \phi(x')$ を高次元空間の特徴ベクトルとして

$$k(x, x') = \phi(x)^T \phi(x')$$

と表される。本研究では、カーネル関数としては、Gaussカーネル

$$k(x, x') = \exp(-\beta \|x - x'\|^2), \quad \beta > 0$$

を用いた。ここで、 β は非線形性を調整するパラメータ (本研究ではカーネルパラメータと呼ぶ) の一種と考えられ、あらかじめ適当に決めておく。

2.3 正則化

一般にモデルの次元が高くなると、関数の表現能力が指数関数的に増大するため汎化能力が落ちる。対してカーネル法では、高次元に保ったまま関数の表現を抑える、正則化という方法を用いる。正則化は、サンプルに対する誤差のほかに負荷項を付け加えた

$$R_{k,\lambda}(\alpha) = (y - K\alpha)^T (y - K\alpha) + \lambda \alpha^T K \alpha, \quad \lambda > 0$$

を最小化することによって、カーネル関数の表現能力を落とすという方法である。正則化の際に加えた $\lambda \alpha^T K \alpha$ を正則化項と呼び、ここでは、その強さを調節している λ を正則化パラメータと呼ぶ。

2.4 カーネル回帰を用いた判別

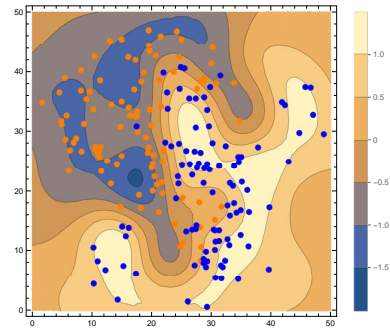
2群データのそれぞれのラベルを $1, -1$ として、カーネル回帰を用いて回帰曲面をつなぐ。 $(1, -1)$ の中点である 0 での等高線を判別曲線とする。カーネル関数のカーネルパラメータ β は非線形性の強度、正則化パラメータ λ はカーネル曲面の平滑度に対応する。

3 提案手法

本研究ではカーネルパラメータ β を固定した時、最適な正則化パラメータ λ を以下の方法で推定する。

$\lambda = 0.001$ から 0.1 までの 0.001 刻みの 100 個の λ の値に対してカーネル回帰をそれぞれ行い、得られた回帰曲面の値の λ に対する変化率を調べ、その最頻値を与えるような λ を最適な正則化パラメータと定める。

ここで曲面値とは、カーネル回帰を行い、得られた回帰曲面の値のことである。曲面値を視覚的に表したのが以下の図で、図中の 0 での等高線を判別曲線とする。



また、 $\lambda = 0.001$ から 0.1 としたのは、正則化パラメータは小さくなるほどオーバーフィッティングしやすいという性質から、過学習にならない程度になるべく小さな λ が、適度に特徴を捉えた最適な判別曲線になると考えられるためである。

4 実験

4.1 実験 1

参考文献 [1] で判別に使われているデータに、本研究の手法を適応した。また今回は、 $\beta = 0.015$ とした。

$\lambda = 0.001$ から 0.1 までの 0.001 刻みの λ に対して回帰曲面の値の λ に対する変化率を求めた結果が以下である。

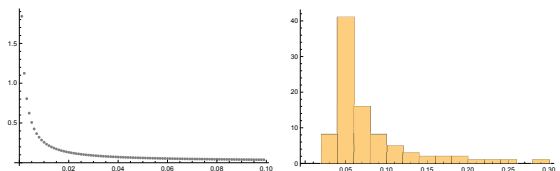


図 1: 回帰曲面の値の λ に対する変化率
図 2: 変化率のヒストグラム

図 2 のヒストグラムより、曲面値の変化量は 0.05 が

一番多いことがわかる。これに最も近い曲面値をとる λ の値は $\lambda = 0.65$ であり、これを λ に推定値とする。
この時のカーネル回帰曲線が以下の図 3 である。

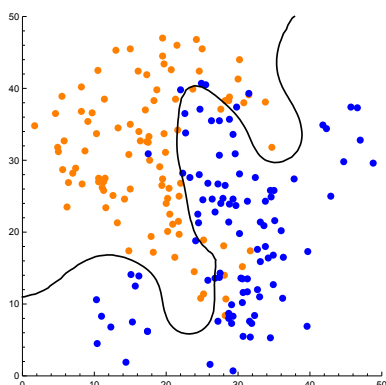


図 3: 判別曲線

4.2 実験 2

参考文献 [2] で判別に使われているデータに、本研究の手法を適応した。また今回は、 $\beta = 0.005$ とした。

実験 1 と同様にして、回帰曲面の値の λ に対する変化率を求めた結果が以下である。

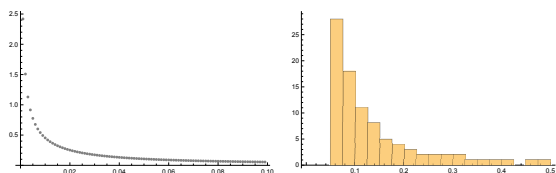
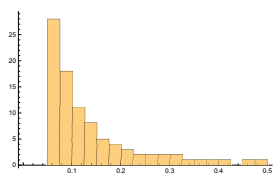


図 4: 回帰曲面の値の λ に対する変化率



ラム

図 5 のヒストグラムより、曲面値の変化量は 0.075 が一番多く、これに最も近い曲面値の値をとる λ の値を調べることにより、 λ を 0.071 と推定値とする。

この時のカーネル回帰曲線が以下の図 6 である。

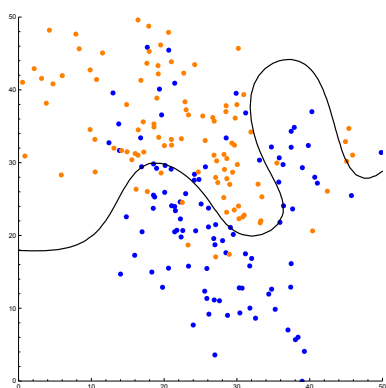
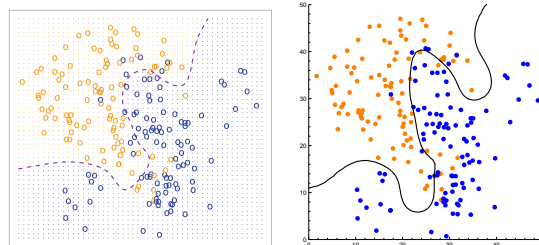


図 6: 判別曲線

5 実験結果のまとめ

推定できたパラメータを用いた判別曲線と、参考文献で行われたベイズ推定による判別曲線をそれぞれ比べた結果が以下である。

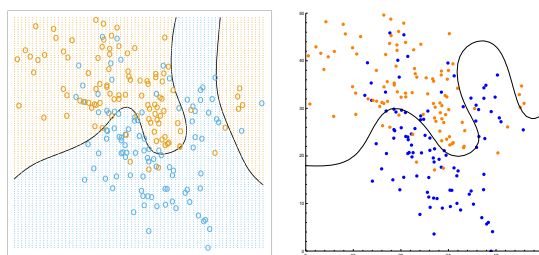


(a) ベイズ推定

(b) 本研究の手法

図 7: 実験 1

実験 1 では参考文献にあるベイズ推定と近い判別曲線が推定できた。本研究の手法の方がより孤立したデータを拾う傾向にあることがわかる。



(a) ベイズ推定

(b) 本研究の手法

図 8: 実験 2

実験 2 ではベイズ推定とは異なる曲線が得られた。また、誤判別率は以下ようになった。どちらの実験でもオーバーフィッティングすることなく、ベイズ推定とほぼ同じ精度で推定できていることがわかる。

	実験1	実験2
ベイズ推定	9.09%	13.0%
本研究の手法	8.59%	14.0%

6 今後の課題

ある推定方法が常に適切ということではなく、どの推定方法が解析するデータに対して適しているかの評価の検討を行いたい。

また、カーネルパラメータの値によっても曲線がかなり異なるため、カーネルパラメータの値の決定が今後の課題である。

参考文献

- [1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Springer, New York, 2013.
- [2] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York, 2009.