

## 4.3. 線形判別機

- 線形判別分析(LDA) ... 2次元以上のデータを各1次元に投影する分類方法。
- 線形判別関数 ...
- 2次判別関数(QDA) ... 変数間で分散が異なる際にも利用できる分類方法。

### 4.3.1. 正則化判別分析

LDAとQDAの折衷案的手法もある。

QDAの各共分散をLDAの共分散に近づける手法である。

$\hat{\Sigma}$ はLDAの合併共分散行列である。

$$\hat{\Sigma}_k(\gamma) = \gamma \hat{\Sigma} + (1 - \gamma) \hat{\sigma}^2 I \quad (4.14)$$

### 4.3.2. 線形判別分析の計算

LDA、QDAの計算には以下の計算が必要である。

$$(x - \hat{\mu}_k)^T \hat{\Sigma}^{-1} (x - \hat{\mu}_k) = [U_k^T (x - \hat{\mu}_k)]^T D_k^{-1} [U_k^T (x - \hat{\mu}_k)]$$

$$\log |\hat{\Sigma}_k| = \sum_l \log d_{kl}$$

### 4.3.3. 階数低減型線形判別分析

正準変量 / 判別変量 ...

- クラス内共分散行列 ...
- クラス間共分散行列 ...
- レイリー商 ...
- 判別座標軸 ...

## 4.4. ロジスティック回帰

ロジスティック回帰は2値の分類問題に適しており、解釈容易性の高さ、過学習のリスクの低さ、計算効率の高さがメリットである。

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i} \quad (i = 1, \dots, n)$$

### 4.4.1. 当てはめ

ロジスティック回帰では、 $X$ が与えられた元での $G$ の条件付き尤度を元に、最尤法で当てはめられるのが一般的。

- 尤度 ... ある前提から結果が得られた時、**逆に前提条件を推測する尤(もっと)もらしさを示す数値**
- スコア方程式  $= \sum_{i=1}^N \{y_i \beta^T x_i - \log(1 + e^{\beta^T x_i})\}$
- 反復重み付け最小2乗法(IRLS) ... 最小2乗法で使っていた残差平方和に重み付けを行う手法。外れ値に強いメリットがある。

### 4.4.3. 2次近似と2次推測

- ラオのスコア検定 ... 項を含んでいるかを検定する。
- ワルド検定 ... 項の除外ができていないかを検定する。

上の2つの検定を繰り返すことが重み付き最小2乗当てはめに等価である？

#### 4.4.4. L1正則化付きロジスティック回帰

L1罰則を付け加えることで、変数選択と係数縮小が可能。

$$\max_{\beta_0, \beta} \left\{ \sum_{i=1}^N [y_i(\beta_0 + \beta^T x_i) - \log(1 + \exp(\beta_0 + \beta^T x_i))] - \lambda \sum_{j=1}^p |\beta_j| \right\} \quad (4.31)$$



## 4.4.5. ロジスティック回帰or線形判別分析

- 同時密度  $Pr(X, G = k) = Pr(X)Pr(G = k|X)$

## 5.1. 導入

これまでの線形的手法も有用だが、常に**真の関数 $f(X)$ が線形であるとは限らない**。一般に、 $f(X)$ は非線形で非加法的だが、**扱いにくいため**線形モデルを利用している。また、入力 $N$ が小さいときや次元 $p$ が大きいときは、線形モデルが唯一過学習なく利用できることは確かである。

- 基底変換 ... 新たな平面を作り、その中で元のベクトルを表現する操作
- 区分的多項式 ... 複数の区間ごとに多項式によって定義される関数
- スプライン ... 区分的に多項式を当てはめて、**接点でなめらかに接続したもの**。
- ウェーブレット関数 ... ウェーブレット変換という、周波解析法に使われる手法で、基底関数に利用されるもの。時間領域の情報を損失しなかったり、 $O(N)$ で計算できたり等のメリットがある。



## 5.2. 区分的多項式(piecewise polynomial) / スプライン(spline)

3次スプラインがよく使われる。これは、過適合を防いだり、そもそも意味がなかったりするため。

$S_0(x), S_1(x), \dots, S_{n-1}(x)$  の  $n$  個の多項式がある時、スプライン曲線を求めるには、

1. 条件式をまとめて行列の式にして、逆行列を使って解く。
2. 各係数の具体的な形を代入と式変形を繰り返して求めて解く。(  $c_j$  を求めるのが大変)

スプライン補間について (その2) #数学 - Qiita

<https://qiita.com/khidaka/items/84610cd890ecb8443d96>

- 回帰スプライン ... スプラインの次数と接点の数と配置を選択する必要がある。
- Bスプライン ... 複数の多項式をなめらかに接続して、1つの基底関数を構成する。



### 5.2.1. 3次自然スプライン(natural cubic spline)

スプラインにおいて、外挿や境界接点を横切る当てはめは、問題が発生する。これを解決する手法に、**3次自然スプライン**がある。元のスプラインと比べて、

- 関数が境界接点上で線形である

という制約が加えられ、自由度が4つ減ることがわかる。

K個の接点を持つ3次自然スプラインは、K個の基底関数を元に表現できる。



## 5.3. フィルタリングと特徴抽出

高次の特徴を持つ際の前処理として、フィルタリングや特徴抽出を行う。

- 音素解析の際には、 $p \times M$ 基底行列 $H$ を構築し、特徴 $x$ から新たな特徴 $x^* = H^T x$ への変換を行う。(フィルタリング)  
この際には、線形ロジスティクス回帰を行った。

この、 $x^* = H^T x$  がウェーブレット変換であり、NNの入力などにも用いられる。

ウェーブレットは、離散値の急変や境界を捉えるために有用であり、NNでは目的変数の予測のために特徴の非線形関数を構築する強力な道具である。





## 5.4. 平滑化スプライン(smoothing spline)

平滑化スプラインは、"元のデータへの近似度"+"関数の滑らかさ"を最適化して、スプライン関数を決定する方法である。(5.9)

関数の滑らかさについては、全ての関数 $f(x)$ における2次導関数とパラメータ $\lambda$ を使って調整する。

$$RSS(f, \lambda) = \sum_{i=1}^N \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\}^2 dt \quad (5.9)$$

動かして学ぶバイオメカニクス#5 ～フィルタリングと加速度計算～ | SPORTS

SENSINGスポーツ科学研究室

[https://note.com/ss\\_sports\\_lab/n/n28b75abefc86#4f7a8f21-0b70-49b8-ab06-873ea2944879](https://note.com/ss_sports_lab/n/n28b75abefc86#4f7a8f21-0b70-49b8-ab06-873ea2944879)



## 5.4.1. 自由度と平滑化

平滑化とは、

線形平滑化とは、推定値が  $\hat{f}(x) = N(N^T N + \lambda \Omega_N)^{-1} N^T y = S_\lambda y$  のように線形になるような平滑化である。

このときの線形演算子  $S_\lambda$  は、**平滑化行列** として知られる。

平滑化スプラインも線形平滑化の例である。

- ハット行列 ...  $H = X(X^T X)^{-1} X^T$
- ラインシュ行列 ...
- デムラー=ラインシュ基底 ...
- 等価カーネル(平滑化行列) ... 訓練データの目標値の線型結合から予測を行う方法。特徴は、共分散行列であること、全データ  $x$  に対してカーネル関数の総和は1であること、等価カーネルは内積で表現できること。

