カーネル法の基礎 4.1節

参考文献

• 金森, 統計的学習理論, 講談社, 2015.

Table of contents

• 線形モデルを用いた学習

線形モデルを用いた学習

判別,回帰: 入力データ(x,y)から入出力の間の関数を学習する (例) 線形モデル

$$\mathcal{M} = \{f(x) = eta^{ ext{T}} \phi(x) \, | \, eta \in \mathbb{R}^D \}$$

ここで、

$$\phi(x) = (\phi_1(x), \ldots, \phi_D(x))^{\mathrm{T}} \in \mathbb{R}^D$$

は入力空間 \mathcal{X} から \mathbb{R}^D への写像

- 線形モデルは,基底関数 $\phi_1(x),\ldots,\phi_D(x)$ の線形結合で関数を表す
- データから線形モデルのパラメータetaを推定量 \hat{eta} で推定
- $\hat{f}(x) = \hat{eta}^{\mathrm{T}} \phi(x)$ を予測に利用

- 回帰問題
 - \circ 入力xにおける出力yの値を $\hat{f}(x)$ で予測
- 2值判別問題
 - \circ $\hat{f}(x)$ の正負で2値ラベルを予測

線形モデルの表現力は、基底関数の数に依存

- 次元Dが大きい
 - *M*の表現力は高くなる
 - 。 パラメータβを推定するための計算量が大きくなる

(例)線形モデルの計算量

観測データ

$$(x_1,y_1),\ldots,(x_n,y_n)\in\mathcal{X} imes\mathbb{R}$$

以下の $n \times D$ 行列Xとn次元ベクトルYを定義.

$$X = (\phi(x_1), \ldots, \phi(x_n))^{\mathrm{T}} \in \mathbb{R}^{n imes D}, \; Y = (y_1, \ldots, y_n)^{\mathrm{T}} \in \mathbb{R}^n$$

 $\leadsto y_i$ を $eta^{\mathrm{T}}\phi(x_i)$ で近似するため,正則化 (正則化項: $\lambda \|eta\|^2$).

$$\min_{eta \in \mathbb{R}} \sum_{i=1}^n (y_i - eta^{\mathrm{T}} \phi(x_i))^2 + \lambda \|eta\|^2 ~~(4.1)$$

を最小にするように学習.

$$\hat{eta} = (X^{\mathrm{T}}X + \lambda I_D)^{-1}X^{\mathrm{T}}Y$$

- $oldsymbol{\hat{eta}}$ を解くために,D次元線形方程式を解く必要がある.
 - \circ Dが非常に大きいと困難.
 - \circ $D=\infty$ では表現力は大きいが,正規方程式を数値的に解けない
- ightsquigarray 高次元モデルを用いた推定量の効率的な計算法の考察,カーネル法との関連の説明 (4.1) の目的関数でパラメータ β に関連する項は...
 - 内積 $\beta^{\mathrm{T}}\phi(x)$
 - $J \mathcal{L} \Delta \|\beta\|^2$

ここで,以下の \mathbb{R}^D の部分空間を考える.

- $S = \operatorname{Span}\{\phi(x_1), \ldots, \phi(x_n)\}$: $\phi(x_1), \ldots, \phi(x_n)$ で張られる空間
- $S^{\perp}=\{v\in\mathbb{R}^D\,|\,v^{\mathrm{T}}w=0\ orall w\in S\}$: Sの直交補空間

射影定理より,

$$eta = eta_S + eta_{S^\perp}, \;\; eta_S \in S, \; eta_{S^\perp} \in S^\perp$$

と一意に直交分解できる.

Sの定義より、以下が成立する.

$$ullet \; eta_{S^{\perp}}^{\mathrm{T}} \phi(x_i) = 0 \; orall i = 1, \ldots, n$$

$$oldsymbol{eta} oldsymbol{eta}^{\mathrm{T}} \phi(x_i) = (eta_S + eta_{S^\perp})^{\mathrm{T}} \phi(x_i) = eta_S^{\mathrm{T}} \phi(x_i)$$

•
$$\|\beta\|^2 = \|\beta_S + \beta_{S^{\perp}}\|^2 = \|\beta_S\|^2 + \|\beta_{S^{\perp}}\|^2$$

(4.1)の目的関数は、以下のように書き直される.

$$L = \sum_{i=1}^n (y_i - eta_S^{ ext{T}} \phi(x_i))^2 + \lambda \|eta_S\|^2 + \lambda \|eta_{S^\perp}\|^2$$

よって,最適解は $eta_{S^\perp}=\mathbf{0}$,すなわち $eta\in S$ において達成される ($eta_{S^\perp}
eq \mathbf{0}$ とすると,第3項のノルムの分で目的関数値が大きくなる).

 \leadsto 最適解の候補として $\phi(x_i)$ の線型結合で表されるパラメータを考えれば十分.

$$eta = \sum_{i=1}^n lpha_i \phi(x_i)$$

とする. さらに, 関数k(x,x')を

$$k(x, x') = \phi(x)^{\mathrm{T}} \phi(x')$$
 (4.2)

とおき,n imes n行列Kを $K=(K_{ij})$, $K_{ij}=k(x_i,x_j)$ とおく.

•
$$eta^{\mathrm{T}}\phi(x_i) = \left(\sum_{j=1}^n lpha_j\phi(x_j)
ight)^{\mathrm{T}}\phi(x_i) = \sum_{j=1}^n lpha_j k(x_j,x_i) = \sum_{j=1}^n K_{ij}lpha_j$$

•
$$\|\beta\|^2 = \left(\sum_{i=1}^n \alpha_i \phi(x_i)\right)^{\mathrm{T}} \left(\sum_{j=1}^n \alpha_j \phi(x_j)\right) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K_{ij}$$

(4.1)の目的関数は

$$L = \sum_{i=1}^n \left(y_i - \sum_{j=1}^n K_{ij}lpha_j
ight)^2 + \lambda \sum_{i=1}^n \sum_{j=1}^n lpha_ilpha_j K_{ij}$$

となる.ここで, $lpha=(lpha_1,\ldots,lpha_n)^{\mathrm{T}}$ とし, $K^{(i)}=(K_{i1},\ldots,K_{in})$ とする.

$$ullet \sum_{j=1}^n K_{ij} lpha_j = \left(egin{array}{ccc} K_{i1} & \cdots & K_{in} \end{array}
ight) \left(egin{array}{c} lpha_1 \ dots \ lpha_n \end{array}
ight) = K^{(i)} lpha$$

$$egin{aligned} ullet \sum_{i=1}^n (y_i - K^{(i)} lpha) &= \left\| \left(egin{array}{ccc} y_1 \ dots \ y_n \end{array}
ight) - \left(egin{array}{ccc} K_{11} & \cdots & K_{1n} \ dots & dots \ K_{n1} & \cdots & K_{nn} \end{array}
ight) \left(egin{array}{ccc} lpha_1 \ dots \ lpha_n \end{array}
ight)
ight\|^2 \ &= \|Y - K lpha\|^2 \end{aligned}$$

•
$$\sum_{i=1}^n \sum_{j=1}^n lpha_i lpha_j K_{ij} = lpha^{\mathrm{T}} K lpha$$

より、(4.1)の目的関数は次のようにベクトルと行列を使って書き直される.

$$L = \|Y - Klpha\|^2 + lpha^{\mathrm{T}} Klpha$$

これはlphaに関して凸である.ここで, $K^{
m T}=K$ に注意すると,

$$rac{\partial L}{\partial lpha} = KY - K(K + \lambda I_n)lpha = \mathbf{0}$$

となる $\alpha = \hat{\alpha}$ で目的関数値が最小になる.

Kが正則だと仮定すると,

$$\hat{lpha} = (K + \lambda I_n)^{-1} Y$$

となる.

よって、回帰関数の推定量として

$$\hat{f}(x) = \hat{eta}^{ ext{T}} \phi(x) = \sum_{i=1}^n \hat{lpha}_i \phi(x_i)^{ ext{T}} \phi(x) = \sum_{i=1}^n \hat{lpha}_i k(x_i, x) ~~(4.3)$$

が得られる.

以上より,次のことが分かる.

- 回帰関数 $\hat{f}(x)$ を求めるためには関数k(x,x')が計算できれば十分、 \circ 基底関数 $\phi(x)$ を求める必要がない.
- 行列 $K\in\mathbb{R}^{n\times n}$ が与えられれば, \hat{lpha} の計算は $\phi(x)$ の次元Dに依存せずにn次線形方程式に帰着される.
- → この考え方はカーネル法として一般化される.