

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

$$\mathcal{M} = \{f(x) = \beta^T \phi(x) \mid \beta \in \mathbb{R}^D\}$$

ここで，

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

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 $\phi_1(x), \dots, \phi_D(x)$ 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 β を推定量 $\hat{\beta}$ で推定
- $\hat{f}(x) = \hat{\beta}^T \phi(x)$ を予測に利用

- 回帰問題
 - 入力 x における出力 y の値を $\hat{f}(x)$ で予測
- 2値判別問題
 - $\hat{f}(x)$ 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
 - \mathcal{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
 - $(x_1, y_1), \dots, (x_n, y_n) \in \mathcal{X} \times \mathbb{R}$

以下の $n \times D$ 行列 X と n 次元ベクトル Y を定義.

$$X = (\phi(x_1), \dots, \phi(x_n))^T \in \mathbb{R}^{n \times D}, \quad Y = (y_1, \dots, y_n)^T \in \mathbb{R}^n$$

$\rightsquigarrow y_i$ を $\beta^T \phi(x_i)$ で近似するため, 正則化 (正則化項: $\lambda \|\beta\|^2$).

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

を最小にするように学習.

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

- $\hat{\beta}$ を解くために、 D 次元線形方程式を解く必要がある。
 - D が非常に大きいと困難。
 - $D = \infty$ では表現力は大きいですが、正規方程式を数値的に解けない

⇨ 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 $\beta^T \phi(x)$
- ノルム $\|\beta\|^2$

ここで，以下の \mathbb{R}^D の部分空間を考える．

- $S = \text{Span}\{\phi(x_1), \dots, \phi(x_n)\}$: $\phi(x_1), \dots, \phi(x_n)$ で張られる空間
- $S^\perp = \{v \in \mathbb{R}^D \mid v^T w = 0 \ \forall w \in S\}$: S の直交補空間

射影定理より,

$$\beta = \beta_S + \beta_{S^\perp}, \quad \beta_S \in S, \quad \beta_{S^\perp} \in S^\perp$$

と一意に直交分解できる.

S の定義より, 以下が成立する.

- $\beta_{S^\perp}^\top \phi(x_i) = 0 \quad \forall i = 1, \dots, n$
- $\beta^\top \phi(x_i) = (\beta_S + \beta_{S^\perp})^\top \phi(x_i) = \beta_S^\top \phi(x_i)$
- $\|\beta\|^2 = \|\beta_S + \beta_{S^\perp}\|^2 = \|\beta_S\|^2 + \|\beta_{S^\perp}\|^2$

(4.1)の目的関数は, 以下のように書き直される.

$$L = \sum_{i=1}^n (y_i - \beta_S^\top \phi(x_i))^2 + \lambda \|\beta_S\|^2 + \lambda \|\beta_{S^\perp}\|^2$$

よって、最適解は $\beta_{S^\perp} = \mathbf{0}$ ，すなわち $\beta \in S$ において達成される ($\beta_{S^\perp} \neq \mathbf{0}$ 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)。

⇨ 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。

$$\beta = \sum_{i=1}^n \alpha_i \phi(x_i)$$

とする。さらに，関数 $k(x, x')$ を

$$k(x, x') = \phi(x)^\top \phi(x') \quad (4.2)$$

とおき， $n \times n$ 行列 K を $K = (K_{ij})$ ， $K_{ij} = k(x_i, x_j)$ とおく。

- $\beta^\top \phi(x_i) = \left(\sum_{j=1}^n \alpha_j \phi(x_j) \right)^\top \phi(x_i) = \sum_{j=1}^n \alpha_j k(x_j, x_i) = \sum_{j=1}^n K_{ij} \alpha_j$
- $\|\beta\|^2 = \left(\sum_{i=1}^n \alpha_i \phi(x_i) \right)^\top \left(\sum_{j=1}^n \alpha_j \phi(x_j) \right) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K_{ij}$

(4.1)の目的関数は

$$L = \sum_{i=1}^n \left(y_i - \sum_{j=1}^n K_{ij} \alpha_j \right)^2 + \lambda \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K_{ij}$$

となる．ここで， $\alpha = (\alpha_1, \dots, \alpha_n)^T$ とし， $K^{(i)} = (K_{i1}, \dots, K_{in})$ とする．

$$\begin{aligned} \bullet \sum_{j=1}^n K_{ij} \alpha_j &= \begin{pmatrix} K_{i1} & \cdots & K_{in} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = K^{(i)} \alpha \\ \bullet \sum_{i=1}^n (y_i - K^{(i)} \alpha) &= \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} K_{11} & \cdots & K_{1n} \\ \vdots & & \vdots \\ K_{n1} & \cdots & K_{nn} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \right\|^2 \\ &= \|Y - K\alpha\|^2 \end{aligned}$$

- $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K_{ij} = \alpha^T K \alpha$

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．

$$L = \|Y - K\alpha\|^2 + \alpha^T K \alpha$$

これは α に関して凸である．ここで， $K^T = K$ に注意すると，

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

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，

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

となる．

よって、回帰関数の推定量として

$$\hat{f}(x) = \hat{\beta}^T \phi(x) = \sum_{i=1}^n \hat{\alpha}_i \phi(x_i)^T \phi(x) = \sum_{i=1}^n \hat{\alpha}_i k(x_i, x) \quad (4.3)$$

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．
 - 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， $\hat{\alpha}$ 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．