

9 Linear Predictors

2016 年 6 月 3 日

1 線形モデル

定義 1 (アフィン関数).

$$L_d = \{h_{w,b} : w \in \mathbb{R}^d, b \in \mathbb{R}\}, \quad (1)$$

where,

$$h_{w,b}(x) = \langle w, x \rangle + b = \sum_{i=1}^d \omega_i x_i + b. \quad (2)$$

$w' = (b, \omega_1, \omega_2, \omega_3, \dots, \omega_d), x' = (1, x_1, x_2, x_3, \dots, x_d)$ をすることによってアフィン関数を斉次関数に変換することができる。

2 線形分類

$$\begin{aligned} \mathcal{X} &= \mathbb{R}^d, \mathcal{Y} = \{-1, +1\} \\ HS_d &= \text{sign} \circ L_d = \{x \rightarrow \text{sign}(h_{w,b}(x)) : h_{w,b} \in L_d\} \end{aligned} \quad (3)$$

2.1 線型計画法

$$\begin{aligned} &\max_{w \in \mathbb{R}^d} \langle u, w \rangle \\ &\text{subject to } Aw \geq v \end{aligned} \quad (4)$$

線型計画 (4) を効率的に解ける手法がある.

線形分類問題が以下の線型計画問題を用いて解けることができる.

$$\begin{aligned} &\max_{w \in \mathbb{R}^d} \langle 0, w \rangle \\ &\text{subject to } (y_i x_{i,j}) w \geq v \end{aligned} \quad (5)$$

Algorithm 1 Batch Perceptron

Input: A training set $(x_1, y_1), \dots, (x_m, y_m)$ **Initialize:** $w^{(1)} = (0, \dots, 0)$

```
1: for  $t = 1, 2, \dots$  do
2:   if  $(\exists i. t. y_i \langle w^{(t)}, x_i \rangle \leq 0)$  then
3:      $w^{(t+1)} = w^{(t)} + y_i x_i$ 
4:   else
5:     output  $w^{(t)}$ 
6:   end if
7: end for
```

2.2 Perceptron

一回の反復には w を $y_i x_i$ 修正する. データセットが分割可能であればアルゴリズムは $(RB)^2$ 以下の反復で終了する. B はデータセットを二分できる w のモジュラの最小値で、 R は $\|x\|$ の最大値である.

$$1 \geq \frac{\langle w^*, w^{(T+1)} \rangle}{\|w^*\| \|w^{(T+1)}\|} \geq \frac{\sqrt{T}}{(RB)^2} \quad (6)$$

まず数学的帰納法を用いて $\langle w^*, w^{(T+1)} \rangle \geq T$ を証明する. 反復 1 : $w^{(1)} = (0, \dots, 0)$, $\langle w^*, w^{(1)} \rangle = 0$ 反復 t :

$$\begin{aligned} \langle w^*, w^{(t+1)} \rangle - \langle w^*, w^{(t)} \rangle &= \langle w^*, w^{(t+1)} - w^{(t)} \rangle \\ &= \langle w^*, y_i x_i \rangle = y_i \langle w^*, x_i \rangle \\ &\geq 1 \end{aligned} \quad (7)$$

次は $\|w^{(T+1)}\|$ の上限を計算する.

$$\begin{aligned} \|w^{(T+1)}\|^2 &= \|w^{(T)} + y_i x_i\|^2 \\ &= \|w^{(T)}\|^2 + 2y_i \langle w^*, x_i \rangle + y_i^2 \|x_i\|^2 \\ &\leq \|w^{(T)}\|^2 + R^2 \leq TR^2 \end{aligned} \quad (8)$$

2.3 VC Dimension

halfspaces の VC Dimension は d である.

1, HS がシャッターできるサイズ d のデータセットが存在する: 線形区間の基底

2, データセットのサイズが d より大きい場合: $\sum_{i \in I} a_i x_i = \sum_{j \in J} |a_j| x_j$

$$0 < \sum_{i \in I} a_i \langle x_i, w \rangle = \langle \sum_{i \in I} a_i x_i, w \rangle = \langle \sum_{j \in J} |a_j| x_j, w \rangle = \sum_{j \in J} |a_j| \langle x_j, w \rangle < 0$$

3 線形回帰

$$\mathcal{H}_{reg} = L_d = \{x \rightarrow \langle w, x \rangle + b : w \in \mathbb{R}^d, b \in \mathbb{R}\} \quad (9)$$

二乗ロス:

$$\ell(h, (x, y)) = (h(x) - y)^2 \quad (10)$$

平均二乗ロス:

$$L_s(h) = \frac{1}{m} \sum_1^m (h(x) - y)^2 \quad (11)$$

3.1 最小二乗法

$$\underset{w}{\operatorname{argmin}} L_s(h_w) = \underset{w}{\operatorname{argmin}} \frac{1}{m} \sum_1^m (\langle w, x_i \rangle - y)^2 \quad (12)$$

$$\frac{2}{m} \sum_1^m (\langle w, x_i \rangle - y) x_i = 0 \quad (13)$$

$$w = A^{-1}b, w = A^+b$$

$$A = \left(\sum_1^m x_i x_i^T \right) \text{ and } b = \sum_i^m y_i x_i \quad (14)$$

3.2 多項式回帰

$$x = p(x) = w_0 + w_1 x + \dots + w_n x^n = \langle w, (1, x, \dots, x^n) \rangle = \langle w, \psi(x) \rangle \quad (15)$$

4 ロジスティック回帰

$$\mathbb{R} \rightarrow [0, 1], \phi_{sig}(z) = \frac{1}{1 + \exp(-z)}, \mathcal{H}_{sig} = \phi_{sig} \circ L_d$$

ロジスティック回帰を用いて二値分類を解くことが多い, $h(x) = \operatorname{sign}(h_w(x) - 1/2)$
 $y = 1$ のときは $1 - h_w(x)$ を小さくしたい, $y = -1$ のときは $h_w(x)$ を小さくしたい

$$1 - h_w(x) = 1 - \frac{1}{1 + \exp(-\langle w, x \rangle)} = \frac{\exp(-\langle w, x \rangle)}{1 + \exp(-\langle w, x \rangle)} = \frac{1}{1 + \exp(\langle w, x \rangle)} = \frac{1}{1 + \exp(y \langle w, x \rangle)} \quad (16)$$

$$h_w(x) = 1 - \frac{1}{1 + \exp(-\langle w, x \rangle)} = \frac{1}{1 + \exp(y \langle w, x \rangle)} \quad (17)$$

$$\ell(h_w(x)) = \frac{1}{1 + \exp(y \langle w, x \rangle)} \quad (18)$$