

## Python を用いた初心者向け AI 実践講座 (中級編) 10/19 配布資料

### 1-2 重回帰と最小二乗法

入力を  $D$  次元ベクトル  $\vec{x}^T = (1, x_1, x_2, \dots, x_D)$ , 出力を  $y$  とすると

$$y = w_0 + w_1 x_1 + w_2 x_2 + \dots + w_D x_D = \vec{w}^T \vec{x} \quad (1.2.1)$$

の  $w_0, w_1, \dots, w_D$  を求める問題を重回帰と呼ぶ。式 (1.2.1) を  $N$  個のデータについて並べて、

$$\vec{\hat{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \\ \vdots \\ \hat{y}_N \end{pmatrix} = \begin{pmatrix} \vec{w}^T x_1 \\ \vec{w}^T x_2 \\ \vdots \\ \vec{w}^T x_n \\ \vdots \\ \vec{w}^T x_N \end{pmatrix} = \begin{pmatrix} \vec{x}_1^T \\ \vec{x}_2^T \\ \vdots \\ \vec{x}_n^T \\ \vdots \\ \vec{x}_N^T \end{pmatrix} \vec{w} = X \vec{w} \quad (1.2.2)$$

と表記することができる。(  $X$  は計画行列と呼ばれている。 ) 正解データ  $y_n$  との誤差を考えると、

$$E = \sum_{n=1}^N (y_n - \hat{y}_n)^2 = \sum_{n=1}^N (y_n - \vec{w}^T \vec{x}_n)^2 \quad (1.2.3)$$

この誤差  $E$  は「最小二乗誤差」と呼ばれ最も標準的な誤差関数である。(ほとんどの機械学習のモデルではこういった誤差関数という類の関数を最小にすることを「学習」と呼んでいる。) この誤差関数を  $\vec{w}$  について微分すると、

$$\begin{aligned} \frac{\partial E}{\partial \vec{w}} &= 0 \\ \Rightarrow X^T X \vec{w} &= X^T \vec{y} \end{aligned} \quad (1.2.4)$$

式 (1.2.4) は重回帰モデルの「正規方程式」と呼ばれている。

$X^T X$  が正則なとき上記方程式の解は

$$\vec{w} = (X^T X)^{-1} X^T \vec{y} \quad (1.2.5)$$

となる。これが重回帰における重みの値となる

(1) 式 (1.2.5) の正規方程式を導け。

### 1-3 パーセプトロン

線形識別関数  $f(x) = w^T x$  を用いて、 $f(x) \geq 0$  のとき  $x \in C_1$  (クラス 1)、 $f(x) < 0$  のとき  $x \in C_2$  (クラス 2) のようにする分類問題を考える。パーセプトロンではまず、重回帰と同じようにデータ  $\vec{x}_i$  と重み  $\vec{w}$  の間の線形結合を考え、以下のような計算を行う。

$$z_i = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_D x_D = \vec{w}^T \vec{x}_i \quad (1.2.6)$$

この計算結果  $z_i$  にしたがって、 $i$  番目のデータに対して次のようなラベル付けを行う。(これがパーセプトロンの識別規則である。)

$$f(\vec{x}_i) = \begin{cases} 1 & z_i \geq 0 \\ 0 & z_i < 0 \end{cases} \quad (1.2.7)$$

パーセプトロンの学習規則は  $i$  番目のデータ  $x_i$  を入力したときの出力  $f(x_i)$  に応じて以下のようにになる。

$$\vec{w}_{i+1} = \begin{cases} \vec{w}_i & f(\vec{x}_i) \geq 0 \\ \vec{w}_i + \eta \vec{x}_i & f(\vec{x}_i) < 0 \end{cases} \quad (1.2.8)$$

(注意、片側のクラスに属するデータの符号を反転させるとどちらのクラスのデータも超平面の同じ側に行うことができるので、どちらのクラスに属していても、分類が正しければ  $f(x) \geq 0$ 、間違えていれば  $f(x) < 0$  になる。)

また、パーセプトロンは 2 クラスの学習データが線形分離可能であれば、有限の学習回数  $M$  で収束することが保証されている。(パーセプトロンの収束定理)

(2) パーセプトロンの収束定理

$$M \frac{D^2(\vec{w}^*)\eta}{d(\eta + 2\alpha)} \leq \phi \leq 1 \implies M \leq d \frac{1 + 2\alpha/\eta}{D_{max}^2}$$

を証明せよ。