

最小二乘法 詳解

人工知能概論 補足

$$(11.6) \quad y = f(x) = w^T x$$

Tは転置

1次元出力 (y_i がスカラー) の場合を考える

x_i は本来の入力ベクトルに定数1を加えたベクトル

$$x_i = (1, \bar{x}_{i,1}, \bar{x}_{i,2}, \dots, \bar{x}_{i,d})^T$$

注釈: ベクトルはすべて列ベクトル, 入力は d 次元(特徴量が d 個)

学習データは N 個あり、そのうちの一個を i で表している

重み(係数)ベクトル w は $(d+1)$ 次元のベクトル $(w_0, w_1, \dots, w_d)^T$

だから

$$y_i = w_0 \cdot 1 + w_1 \cdot \bar{x}_{i,1} + w_2 \cdot \bar{x}_{i,2} + \dots + w_d \cdot \bar{x}_{i,d} = \sum_{k=0}^d w_k \bar{x}_{i,k}$$

$$= (w_0, w_1, \dots, w_d) \begin{pmatrix} 1 \\ \bar{x}_{i,1} \\ \vdots \\ \bar{x}_{i,d} \end{pmatrix} = w^T x_i$$

内積による表現

$$(11.7) \mathbf{w}^* = \operatorname{argmin}_{\mathbf{w}} E(\mathbf{w})$$

|| は『ノルム』を表す記号

$$(11.8) = \operatorname{argmin}_{\mathbf{w}} \sum_i \|y_i - \mathbf{w}^T x_i\|^2$$

$E(\mathbf{w})$ は評価関数 (コスト関数とも言う)

この値を最小にする重み $\mathbf{w}$ を求めたい --- その値を $\mathbf{w}^*$ で表す
評価関数として

学習データに対し、「 y_i と $\mathbf{w}^T x_i$ の差の2乗」の総和
を考える。

その最小値を与える $\mathbf{w}$ が求める答え

偏微分の計算 (11.9) & (11.10)

最小値を求める --- 最小値はすくなくとも「極値」

極値では、偏微分が0

$$\begin{aligned} E(w) &= \sum_i \|y_i - w^T x_i\|^2 \\ &= \sum_i (y_i - w^T x_i)^T (y_i - w^T x_i) \end{aligned}$$

y_i が1次元(スカラー)なら
単純に $\sum_i (y_i - w^T x_i)^2$
左の式は多次元でも成立

$\frac{\partial E}{\partial w_j}$ を求める。この偏微分は w_j だけを変数、他は定数とみなして微分

ここで $(y_i - w^T x_i) = (y_i - (w_0 \cdot 1 + w_1 \cdot \bar{x}_{i,1} + w_2 \cdot \bar{x}_{i,2} + \cdots + w_d \cdot \bar{x}_{i,d}))$

だから $\frac{\partial E}{\partial w_j} = -2 \sum_i x_{i,j} (y_i - w^T x_i)$

極値の条件から w^* を求める

$$\frac{\partial E}{\partial w_j} = -2 \sum_i x_{i,j} (y_i - w^T x_i) = 0$$

よって、

$$\sum_i x_{i,j} (y_i - w^T x_i) = 0$$

左辺：

$$\sum_i x_{i,j} (y_i - w^T x_i) = \sum_i x_{i,j} y_i - \sum_i x_{i,j} w^T x_i$$

評価関数 $E(w)$ の最小値を与える w^* は極値の条件を満たすから

$$\sum_i x_{i,j} y_i - \sum_i x_{i,j} w^{*T} x_i = 0$$

(11.12)~(11.15)

$$\sum_i x_{i,j} y_i - \sum_i x_{i,j} w^{*T} x_i = 0$$

これより
$$\sum_i x_{i,j} y_i = \sum_i x_{i,j} w^{*T} x_i \quad (11.12)$$

ここで、 $w^{*T} x_i = w_0^* \cdot 1 + w_1^* \cdot \bar{x}_{i,1} + w_2^* \cdot \bar{x}_{i,2} + \dots + w_d^* \cdot \bar{x}_{i,d} = x_i^T w^*$ だから

$$\sum_i x_{i,j} y_i = \sum_i x_{i,j} x_i^T w^* \quad (11.13)$$

演習問題LSM-1:これから (11.14)および(11.15)を導け

$$(11.14) \quad \mathbf{X}\mathbf{y}^T = (\mathbf{X}\mathbf{X}^T)\mathbf{w}^*$$

$$(11.15) \quad \mathbf{w}^* = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{y}^T$$

ここで $\mathbf{x}_i$ はi番目の学習データ $\mathbf{x}_i = (1, \bar{x}_{i,1}, \bar{x}_{i,2}, \dots, \bar{x}_{i,d})^T$ で、 $\mathbf{y}_i$ はその出力ベクトルとして、 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ とする

演習問題LSM-2

出力 y_i は1次元、入力は定数1を付加した2次元ベクトル $(1, x_i)$ とする

(11.14)式の左辺が $\begin{pmatrix} \sum_i y_i \\ \sum_i x_i y_i \end{pmatrix}$

また右辺が、 $\begin{pmatrix} \sum_i 1 & \sum_i x_i \\ \sum_i x_i & \sum_i x_i^2 \end{pmatrix} w^*$ で表されることを示せ。

またこのことを用いて、演習11-3を解いてみよ。