最小二乗法　詳解

人工知能概論　補足

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

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

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

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$\boldsymbol{w}$は $(d+1)$次元のベクトル$(w_0, w_1, \ldots, w_d)^T$

だから

$$y_i = w_0 \cdot 1 + w_1 \cdot \bar{x}_{i,1} + w_2 \cdot \bar{x}_{i,2} + \cdots + w_d \cdot \bar{x}_{i,d}$$

$$= \sum_{k=0}^{d} w_k \bar{x}_{i,k}$$

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

内積による表現

$$(11.7)\quad w^* = \underset{w}{\operatorname{argmin}}\, E(w)$$

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

‖　‖は『ノルム』を表す記号

$\boldsymbol{E(w)}$ 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$\boldsymbol{w}$ を求めたい ---その値を$\boldsymbol{w}^*$で表す

評価関数として

学習データに対し、「y_iと$w^T x_i$の差の2乗」の総和

を考える。

その最小値を与える$\boldsymbol{w}$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

$$E(w) = \sum_{\mathrm{i}} \| y_{\mathrm{i}} - w^T x_{\mathrm{i}} \|^2$$

$$= \sum_{\mathrm{i}} (y_{\mathrm{i}} - w^T x_{\mathrm{i}})^T (y_{\mathrm{i}} - w^T x_{\mathrm{i}})$$

y_{i}が1次元(スカラー)なら
単純に$\sum_{\mathrm{i}} (y_{\mathrm{i}} - w^T x_{\mathrm{i}})^2$
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 $(y_{\mathrm{i}} - w^T x_{\mathrm{i}}) = (y_{\mathrm{i}} - (w_0 \cdot 1 + w_1 \cdot \bar{x}_{\mathrm{i},1} + w_2 \cdot \bar{x}_{\mathrm{i},2} + \cdots + w_{\mathrm{d}} \cdot \bar{x}_{\mathrm{i},d}))$

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

極値の条件からw*を求める

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

よって、

$$\sum_i x_{i,j}\,(y_\mathrm{i} - w^T x_\mathrm{i}) \quad = 0$$

左辺:

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

評価関数$E(\mathbf{w})$の最小値を与える$\mathbf{w}^*$ は極値の条件を満たすから

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

(11.12)~(11.15)

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

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

ここで、${w^*}^T x_\mathrm{i}$=w_0^*•1+w_1^*•$\bar{x}_{\mathrm{i},1}$ +w_2^*•$\bar{x}_{\mathrm{i},2} + \cdots + w_d^*$•$\bar{x}_{\mathrm{i},d}$=${x_\mathrm{i}}^T w^*$　だから

$\sum_i x_{i,j}\, y_\mathrm{i} = \sum_i x_{i,j} x_\mathrm{i}\, {w^*}^T$　　(11.13)

演習問題LSM-1:これから (11.14)および(11.15)を導け

(11.14)　$\boldsymbol{X}\boldsymbol{y}^\mathrm{T} = (\boldsymbol{X}\boldsymbol{X}^\mathrm{T})w^*$

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

ここで$\boldsymbol{x}_\mathrm{i}$ はi番目の学習データ $\boldsymbol{x}_\mathrm{i} = (1, \bar{x}_{\mathrm{i},1}, \bar{x}_{\mathrm{i},2}, \ldots, \bar{x}_{\mathrm{i},d})^\mathrm{T}$ で、$\boldsymbol{y}_\mathrm{i}$ はその出力ベクトルとして、$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_\mathrm{N})$, $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, .., \boldsymbol{y}_\mathrm{N})$　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。