Assignment of Data Mining (1) April 12, 2022

Name: 階戸　弾　　　ID: 1124525046

1.
A linear regression model of multi-dimensional input $\boldsymbol{x} \in \mathbb{R}^D$ and output $y \in \mathbb{R}$ is defined as

$$f(\boldsymbol{x}, \boldsymbol{w}) = w_0 + \sum_{d=1}^{D} w_d x_d = \boldsymbol{w}^T \boldsymbol{x},$$

where $\boldsymbol{w} = (w_0, w_1, \dots, w_D)^T$ is a parameter vector and $\boldsymbol{x} = (1, x_1, \dots, x_D)^T$ is an input vector. Given N samples $(\boldsymbol{x}_n, y_n), n = 1 \dots, N$, find the optimum $\boldsymbol{w}$ minimizing the squared error:

$$\begin{aligned} J(\boldsymbol{w}) &= \frac{1}{2}\sum_{n=1}^{N}\{f(\boldsymbol{x}_n, \boldsymbol{w}) - y_n\}^2 \\ &= \frac{1}{2}(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y})^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}), \end{aligned}$$

where $\boldsymbol{y} = (y_1, \dots, y_N)^T$ and data matrix

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1^T \\ \boldsymbol{x}_2^T \\ \vdots \\ \boldsymbol{x}_N^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1D} \\ 1 & x_{21} & \cdots & x_{2D} \\ 1 & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{ND} \end{pmatrix}$$

Answer:

$$J(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{f(\boldsymbol{x}_n, \boldsymbol{w}) - y_n\}^2 = \frac{1}{2}(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y})^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y})$$

より，

$$\begin{aligned} \mathrm{J}(\boldsymbol{w}) &= \frac{1}{2}\{(\boldsymbol{X}\boldsymbol{w})^T\boldsymbol{X}\boldsymbol{w} - (\boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}\boldsymbol{y} - \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{w} + \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y}\} \\ &= \frac{1}{2}\{(\boldsymbol{X}\boldsymbol{w})^T\boldsymbol{X}\boldsymbol{w} - ((\boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}\boldsymbol{y})^{\boldsymbol{T}} - (\boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}} + \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y}\} \\ &= \frac{1}{2}\{(\boldsymbol{X}\boldsymbol{w})^T\boldsymbol{X}\boldsymbol{w} - \boldsymbol{2}(\boldsymbol{X}\boldsymbol{w})^{\boldsymbol{T}}\boldsymbol{y} + \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y}\} \\ &= \frac{1}{2}\{\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{X}\boldsymbol{w} - \boldsymbol{2}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{w}^{\boldsymbol{T}}\boldsymbol{y} + \boldsymbol{y}^{\boldsymbol{T}}\boldsymbol{y}\} \end{aligned}$$

ここで最適化パラメータ$\boldsymbol{w}$は，$J(\boldsymbol{w})$を最小とする値であるため，

$$\frac{\vartheta}{\vartheta \boldsymbol{w}} J(\boldsymbol{w}) = \mathit{0}$$

とする．

$$
\begin{aligned}
\frac{\vartheta}{\vartheta \boldsymbol{w}} J(\boldsymbol{w}) &= -\mathbf{X}^T\mathbf{y} + \frac{1}{2}(\mathbf{X}^T\mathbf{X} + (\mathbf{X}^T\mathbf{X})^{\boldsymbol{T}})\boldsymbol{w} \\
&= -\mathbf{X}^T\mathbf{y} + \frac{1}{2}(\mathbf{X}^T\mathbf{X} + \mathbf{X}^T\mathbf{X})\boldsymbol{w} \\
&= -\mathbf{X}^T\mathbf{y} + \mathbf{X}^T\mathbf{X}\boldsymbol{w} \\
&= 0
\end{aligned}
$$

よって，$\mathbf{X}^T\mathbf{X}\boldsymbol{w} = \mathbf{X}^T\mathbf{y}$となるため，多重回帰分析の最適化パラメータは

$$\boldsymbol{w} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

となる．