CREED REILLY
05/31/2021
HW# 1

$s_1 = 150$ cm
$s_2 = 155$ cm
$s_3 = 165$ cm
$s_4 = 190$ cm

1) $\hat{y}_{KNN} = \dfrac{\sum_{i=1}^{k} y_i}{k}$, $k = 3$

$H - S_i$, cm

P	Height (cm)	Weight (kg)	S_1	S_2	S_3	S_4
1	171	80	21	16	6	-19
2	168	78	18	13	3	-22
3	191	100	41	36	26	1
4	182	80	32	27	17	-8
5	150	65	0	-5	-15	-40
6	178	83	28	23	13	-12

$S_1 : \hat{y}_{knn} = \dfrac{80+78+65}{3} = \boxed{74.33 \text{ kg} = \hat{y}_{150}}$

$S_2 : \hat{y}_{knn} = \dfrac{80+78+65}{3} = \boxed{74.33 \text{ kg} = \hat{y}_{155}}$

$S_3 : \hat{y}_{knn} = \dfrac{80+78+83}{3} = \boxed{80.33 \text{ kg} = \hat{y}_{165}}$

$S_4 : \hat{y}_{knn} = \dfrac{100+80+83}{3} = \boxed{87.67 \text{ kg} = \hat{y}_{190}}$

$S_1 = 150$ cm

$S_2 = 155$ cm

$S_3 = 165$ cm

$S_4 = 190$ cm

2) $$\hat{y}_{KNN} = \frac{\sum_{i=1}^{K} w_i y_i}{K}, \quad K = 3$$

$H - S_i$, cm

P	Height (cm)	Weight (kg)	S_1	S_2	S_3	S_4
1	171	80	21	16	6	-19
2	168	78	18	13	3	-22
3	191	100	41	36	26	1
4	182	80	32	27	17	-8
5	150	65	0	-5	-15	-40
6	178	83	28	23	13	-12

$$\hat{y}_{150} = \lim_{n \to 0} \frac{\frac{1}{n}(65) + \frac{1}{18}(78) + \frac{1}{21}(80)}{\frac{1}{n} + \frac{1}{18} + \frac{1}{21}} = \boxed{65 \text{ kg} = \hat{y}_{150}}$$

$$\hat{y}_{155} = \frac{\frac{1}{5}(65) + \frac{1}{13}(78) + \frac{1}{16}(80)}{\frac{1}{5} + \frac{1}{13} + \frac{1}{16}} = \boxed{70.71 \text{ kg} = \hat{y}_{155}}$$

$$\hat{y}_{165} = \frac{80/6 + 78/3 + 83/13}{1/6 + 1/3 + 1/13} = \boxed{79.24 \text{ kg} = \hat{y}_{165}}$$

$$\hat{y}_{190} = \frac{100/1 + 80/8 + 83/12}{1 + 1/8 + 1/12} = \boxed{96.76 \text{ kg} = \hat{y}_{190}}$$

$$3)\quad J(\underline{x}) = \underline{x}^T Q \underline{x} + d^T x + \cancel{c} \leftarrow \nabla_x c = 0$$

throw away

$$= (1\times n)(n\times n)(n\times 1) + (1\times n)(n\times 1)$$

$$= (1\times 1) + (1\times 1) = \text{scalar}$$

$$Q = Q^T \in \mathbb{R}^{n\times n},\quad \underline{x}, \underline{d} \in \mathbb{R}^n,\quad c \in \mathbb{R}^1$$

$$J(\underline{x}) = [x_1 \cdots x_n] \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{12} & q_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ q_{1n} & & & q_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$+ [d_1 \cdots d_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + c$$

$$\begin{array}{cc} (1\times n) & (n\times 1) \\ = \left[\sum_{i=1}^{n} q_{i1} x_i \cdots \sum_{i=1}^{n} q_{in} x_i \right] & \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \sum d_i x_i \end{array}$$

$\longrightarrow$

$$J(\underline{x}) = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} q_{ij} x_i \right) x_j + \sum d_i x_i \rightarrow ①$$

$$= \sum_{i=1}^{n} q_{i1} x_i x_1 + \sum_{j=1}^{n} q_{1j} x_1 x_j + \sum_{j=2}^{n} \sum_{i=2}^{n} q_{ij} x_i x_j$$

symmetry

$$\frac{\partial J}{\partial x_1} = \sum_{i=1}^{n} q_{i1} x_i + \sum_{j=1}^{n} q_{1j} x_j = 2 \sum_{i=1}^{n} q_{1i} x_i \quad \text{since } q_{i1} = q_{1i}$$

$$\frac{\partial J}{\partial x} = \nabla_x (J) = 2 \begin{bmatrix} \sum_{i=1}^{n} q_{1i} x_i \\ \vdots \\ \sum_{i=1}^{n} q_{ni} x_i \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_x \end{bmatrix} = 2Q\underline{x} + \underline{d}$$

$$\boxed{\therefore \nabla_x J(x) = 2Q\underline{x} + \underline{d}}$$

$$\frac{\partial^2}{\partial x \partial x^T} (J) = 2Q\underline{x} \left[\frac{\partial}{\partial x_1} \cdots \frac{\partial}{\partial x_n} \right] + \underline{d} \left[\partial/\partial x_1 \cdots \partial/\partial x_n \right]$$

$$= 2Q \begin{bmatrix} \partial x_1/\partial x_1 & \cdots & \partial x_1/\partial x_n \\ \vdots & \ddots & \vdots \\ \partial x_n/\partial x_1 & \cdots & \partial x_n/\partial x_n \end{bmatrix}$$

$$= 2Q \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix} = 2QI = 2Q$$

$$\boxed{\therefore \frac{\partial^2}{\partial x \partial x^T} J(x) = H = 2Q}$$

4) $\hat{y} = \beta_0 + \beta_1 \hat{x}_1 + \beta_2 \hat{x}_2 + \cdots + \beta_p \hat{x}_p$

$\underline{\hat{x}}' = [1 \; x_1 \cdots x_p]$

$$X_{n \times (p+1)} = \begin{bmatrix} 1 & x_1^1 & x_2^1 & \cdots & x_p^1 \\ & & \vdots & & \\ 1 & x_1^n & x_2^n & \cdots & x_p^n \end{bmatrix}, \quad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\underline{y} = X\underline{\beta} \rightarrow \underline{\beta} = X^{-1}\underline{y}$$

$$\hat{y} = \hat{x}^T \underline{\beta} = \hat{x}^T X^{-1} \underline{y} \qquad (1 \times (p+1))((p+1) \times n)(x \times 1)$$
$$= (1 \times 1)$$

KNN is a method in which a predicted output is a function of k-samples of test outputs, where $1 \leq k \leq n$. Linear regression uses all test outputs to predict a test output, so linear regression is the special test case where $k = n$.

5) $\underline{y} \in \mathbb{R}^n$, $\hat{\underline{y}} = X \underbrace{(X^T X)^{-1} X^T}_{\ell} \underline{y}$

$X \in \mathbb{R}^{n \times (p+1)}$

$\ell = (X^T X)^{-1} X^T$

$\ell \in \mathbb{R}^{[((p+1) \times n)(n \times (p+1))]^{-1} (p+1) \times n}$

$\in \mathbb{R}^{((p+1) \times (p+1)) \times ((p+1) \times n)}$

$\in \mathbb{R}^{(p+1) \times n}$

$\ell_y \in \mathbb{R}^{[(p+1) \times n][n \times 1]}$

$\in \mathbb{R}^{(p+1) \times 1}$ $\leftarrow$ column vector $\underline{m} = \begin{bmatrix} m_1 \\ \vdots \\ m_{p+1} \end{bmatrix}$

where $m_1 \dots m_{p+1} \in \mathbb{R}$

$$\hat{\underline{y}} = X \underline{m} = \begin{bmatrix} | & & | \\ \underline{x}_1 & \cdots & \underline{x}_{p+1} \\ | & & | \end{bmatrix} \begin{bmatrix} m_1 \\ \vdots \\ m_{p+1} \end{bmatrix}$$

$$= m_1 \underline{x}_1 + \cdots + m_{p+1} \underline{x_{p+1}} = \text{span}\{x_1, \cdots, x_{p+1}\}$$

$$= \text{col}(X)$$

$$\boxed{\therefore \hat{y} = X(X^T X)^{-1} X^T \underline{y} \in \text{col}(X)}$$

6) If $\hat{\beta}$ minimizes $RSS(\beta)$ then $\underline{y} - \hat{\underline{y}}$ is $\perp$ to col(X)

$$\hat{\underline{y}} = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n = \underline{x}^T \hat{\underline{\beta}}$$

$$\underline{y} - \hat{\underline{y}} \in N(X) \rightarrow X(\underline{y} - \hat{\underline{y}}) = \underline{0}$$

$X\underline{y} = X\hat{\underline{y}} \leftarrow$ $\hat{y}$ is the mapping of y onto col(X)

$\therefore$ $\hat{\underline{y}} \in$ col(X)

$$RSS(\beta) = \|y - X\beta\|^2 = (\underline{y} - X\underline{\beta})^T(\underline{y} - X\underline{\beta})$$

$$= \frac{\partial}{\partial\beta} RSS(\beta) = \frac{\partial}{\partial\beta}\left(\underline{y}^T\underline{y} - \underline{y}^T X\underline{\beta} - (X\underline{\beta})^T\underline{y} + (X\underline{\beta})^T(X\underline{\beta})\right)$$

$$= -X^T\underline{y} - X^T\underline{y} + 2X^TX\beta = 0$$

$$\cancel{2}X^TX\beta = \cancel{2}X^Ty$$

$$\beta = (X^TX)^{-1}X^Ty$$

$\frac{\partial^2}{\partial\beta^2} RSS(\beta) = 2X^TX \leftarrow$ positive semi definite

$\nwarrow$ thus $RSS(\beta)$ is a <u>min</u>

$\longrightarrow$

$\hat{\underline{y}} = X\hat{\underline{\beta}} = X(X^TX)^{-1}X^T\underline{y}$

from Q5, this $= X\underline{m}$, where $m_i \in \mathbb{R}$

therefore $\hat{\underline{y}} \in col(X)$

$$\boxed{\begin{array}{l}\therefore \text{ when } \beta \text{ is chosen so } ess(\beta) \text{ is a min,} \\ \hat{\underline{y}} \in col(X) \text{ \& therefore } \underline{y}-\underline{y} \in N(X)\text{, or} \\ \text{orthogonal to the } col(X).\end{array}}$$