HW 1

↳ Hardik Prajapati (2678294168)

1. Sample Data:-

Person	Height (cm)	Weight (kg)
1	171	80
2	168	78
3	191	100
4	182	80
5	150	65
6	178	83

Test Data:-

$x_1 = 150\,cm$

$x_2 = 155\,cm$

$x_3 = 165\,cm$

$x_4 = 190\,cm.$

Model:-

$$\hat{y}_{KNN} = \frac{y_1 + y_2 + \cdots + y_k}{k}, \quad \text{Given } k = 3.$$

Prediction:-

i) $x_1 = 150\,cm, \quad y_1 = 65\,kg, \; y_2 = 78\,kg, \; y_3 = 80\,kg.$

$$\Rightarrow \hat{y}_1 = \frac{65 + 78 + 80}{3} = 74.33\,kg.$$

ii) $x_2 = 155\,cm, \quad y_1 = 65\,kg, \; y_2 = 78\,kg, \; 80\,kg$

$$\Rightarrow \hat{y}_2 = \frac{65 + 78 + 80}{3} = 74.33\,kg.$$

iii) $x_3 = 165cm$, $y_1 = 78kg$, $y_2 = 80kg$, $y_3 = 83kg$.

$$\Rightarrow \hat{y}_3 = \frac{78+80+83}{3} = \underline{\underline{80.33kg}}$$

iv) $x_4 = 190cm$, $y_1 = 100kg$, $y_2 = 80kg$, $y_3 = 83kg$.

$$\Rightarrow \hat{y}_4 = \frac{100+80+83}{3} = 87.66kg$$

Ans:- Predictions:-

Person	Height	Predicted Weight (kg)
1	150	74.33
2	155	74.33
3	165	80.33
4	190	87.66

2. **Sample Data:-**

Person	Height (cm)	Weight (kg)
1	171	80
2	168	78
3	191	100
4	182	80
5	150	65
6	178	83

Test Sample:-

$x_1 = 150\text{cm}$

$x_2 = 155\text{cm}$

$x_3 = 165\text{cm}$

$x_4 = 190\text{cm}$

Model:-

$$\hat{y}_{kNN} = \frac{w_1 y_1 + w_2 y_2 + \dots + w_k y_k}{w_1 + w_2 + \dots + w_k}$$

$w_i = \frac{1}{d_i}$, d_i = Distance

Prediction:-

i) $x_1 = 150\text{cm}$, $y_1 = 65\text{kg}$, $d_1 = 0$

$y_2 = 78\text{kg}$, $d_2 = 18$

$y_3 = 80\text{kg}$ $d_3 = 21$

$\Rightarrow \hat{y}_1 = 65\text{kg}$, as $w_1 = \frac{1}{0} = \infty$ dominates over all weights

ii) $x_2 = 155\text{cm}$, $y_1 = 65\text{kg}$, $d_1 = 5$
$y_2 = 78\text{kg}$, $d_2 = 13$
$y_3 = 80\text{kg}$, $d_3 = 16$.

$$\Rightarrow \hat{y}_2 = \frac{\left(\frac{1}{5}\right)(65) + \left(\frac{1}{13}\right)(78) + \left(\frac{1}{16}\right)(80)}{\left(\frac{1}{5} + \frac{1}{13} + \frac{1}{16}\right)}$$

$$= \underline{\underline{70.70\text{ kg}}}.$$

iii) $x_3 = 165\text{cm}$, $y_1 = 78\text{kg}$, $d_1 = 3$
$y_2 = 80\text{kg}$, $d_2 = 6$
$y_3 = 83\text{kg}$, $d_3 = 13$

$$\Rightarrow \hat{y}_3 = \frac{\left(\frac{1}{3}\right)(78) + \left(\frac{1}{6}\right)(80) + \left(\frac{1}{13}\right)(83)}{\left(\frac{1}{3} + \frac{1}{6} + \frac{1}{13}\right)}$$

$$= \underline{\underline{79.24\text{ kg}}}.$$

iv) $x_4 = 190\text{cm}$, $y_1 = 100\text{kg}$, $d_1 = 1$
$y_2 = 80\text{kg}$, $d_2 = 8$
$y_3 = 83\text{kg}$, $d_3 = 12$

$$\Rightarrow \hat{y}_4 = \frac{\left(\frac{1}{1}\right)(100) + \left(\frac{1}{8}\right)(80) + \left(\frac{1}{12}\right)(83)}{\left(\frac{1}{1} + \frac{1}{8} + \frac{1}{12}\right)}$$

$$= \underline{\underline{96.75\text{ kg}}}.$$

Ans:-

Predicted labels:-

Person	Height (cm)	Predicted Weight (kg)
1	150	79.97
2	155	70.70
3	165	79.24
4	190	96.75

3. $J(x) = x^T Q x + d^T x + c$, $Q = Q^T \in R^{n\times n}$
$x, d \in R^n$
$c \in R.$

To show:- $\nabla_x J(x) = 2Qx + d$

$$H = \frac{\partial^2 J}{\partial x \partial x^T} = 2Q \quad ; \; H_{i,j} = \frac{\partial^2 J}{\partial x_i \partial x_j}$$

Proof:-

We notice that $J(x)$ is a scalar quantity as

$$J(x) = \underset{(1\times n)}{x^T} \underset{(n\times n)}{Q} \underset{(n\times 1)}{x} + \underset{(1\times n)}{d^T} \underset{(n\times 1)}{x} + \underset{(1\times 1)}{c}$$

$\Rightarrow \nabla_x J(x)$ is a vector quantity.

Now,

$$J(x) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j q_{ij} + \sum_{i=1}^{n} d_i x_i + c$$

$$\frac{\partial J(x)}{\partial x_k} = \sum_{i=1}^{n} x_i q_{ik} + \sum_{j=1}^{n} x_j q_{kj} + d_k + 0$$

$$\Rightarrow \nabla_x J(x) = Qx + Q^T x + d$$

$$\Rightarrow \nabla_x J(x) = 2Qx + d \quad \rightarrow \{\because Q = Q^T\}$$

Now, $\frac{\partial J}{\partial x_k} = 2\sum_{i=1}^{n} x_i q_{ik} + d_k$

$$\frac{\partial\left(\frac{\partial J}{\partial x_n}\right)}{\partial x_m} = 2q_{mn} + 0$$

$$\Rightarrow H_{ij} = \frac{\partial^2 J}{\partial x_i \partial x_j} = 2q_{ij}$$

$$\Rightarrow H = \frac{\partial^2 J}{\partial x \partial x^T} = 2Q.$$

↳ Hence proved.

4. Test vector :- $x'_{1 \times p}$

Training set label vector :- y

(Augmented) Feature matrix :- $X_{n \times (p+1)}$

$$\hat{y} = X \hat{B}$$

$$\Rightarrow \hat{y} = X (X^T X)^{-1} X^T y$$

$$\Rightarrow \hat{y} = H y$$

↳ We observe that $\hat{y}$ is a Linear combination of Training set label vector y.

↳ For kNN model,

$$\hat{y} = \frac{w_1 y_1 + w_2 y_2 + \cdots w_n y_n}{w_1 + w_2 + w_3 + \cdots w_n}$$

This can also be interpreted as a Linear combination of Training set label vector y, where w_i are the weights.

$\Rightarrow$ Hence, $\hat{y}$ can be viewed as special case of KNN regression, where $\sum_{i=1}^{n} w_i = 1$

5. $y \in \mathbb{R}^n$

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

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

Now,

$$H = X(X^T X)^{-1} X^T$$

$$\Rightarrow H^T = \left(X(X^T X)^{-1} X^T\right)^T$$

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

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

$$\Rightarrow H^T = H.$$

Now,

$$H^2 = \left(X(X^T X)^{-1} X^T\right)\left(X(X^T X)^{-1} X^T\right)$$

$$= X(X^T X)^{-1} \underbrace{(X^T X)(X^T X)^{-1}}_{I} X^T$$

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

$$\Rightarrow H^2 = H.$$

Also, $X^T H = X^T$; $H^T X = X$; $HX = X$

$\Rightarrow H = X(X^T X)^{-1} X^T$ is a projection matrix onto space spanned by $x_0, x_1, \dots x_p$

Now,

For $\hat{y} = Hy$ to have a solution

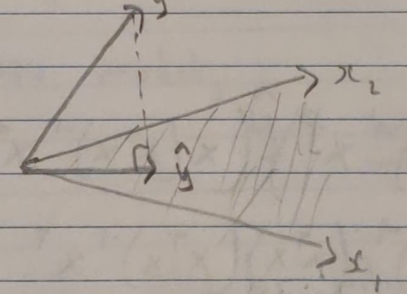
$\Rightarrow \hat{y}$ is a linear combination of $[x_0, x_1 \cdots x_p]$

$\Rightarrow \hat{y}$ lies in column space of X.

6. $RSS(\beta) = ||y - X\beta||^2$

$\text{Column}(X) = \text{span}\{x_1, \cdots, x_p\}$

↳ Let $\hat{y} \in \text{Col}(X)$



↳ Assume $\tilde{y} \in \text{Col}(X)$, $\tilde{y} = X\tilde{\beta}$

Now, $||y - \tilde{y}||^2 = ||y - \hat{y} + \hat{y} - \tilde{y}||^2$

$$= \langle (y - X\hat{\beta}) + (X\hat{\beta} - X\tilde{\beta}), (y - X\hat{\beta}) + (X\hat{\beta} - X\tilde{\beta}) \rangle$$

$$= \langle y - X\hat{\beta}, y - X\hat{\beta} \rangle + \langle y - X\hat{\beta}, X\hat{\beta} - X\tilde{\beta} \rangle + \langle X\hat{\beta} - X\tilde{\beta}, y - X\hat{\beta} \rangle$$

$$+ \langle X\hat{\beta} - X\tilde{\beta}, X\hat{\beta} - X\tilde{\beta} \rangle$$

$$\Rightarrow \|y-\tilde{y}\|^2 = \|y-X\hat{\beta}\|^2 + 2\overbrace{\langle y-X\hat{\beta}, X\hat{\beta}-X\tilde{\beta}\rangle}^{=0} + \|X\hat{\beta}-X\tilde{\beta}\|^2$$

$$\Rightarrow \|y-\tilde{y}\|^2 = \|y-X\hat{\beta}\|^2 + \underbrace{\|X\beta'\|^2}_{\geq 0} \quad \text{where } \beta' = \hat{\beta}-\tilde{\beta}$$

$$\Rightarrow \|y-\tilde{y}\|^2 \geq \|y-\hat{y}\|^2$$

$\Rightarrow \|y-\hat{y}\|^2$ is the minimum RSS

$\Rightarrow$ According to Pythagoras theorem, this is possible only when $(y-\hat{y})$ is the shortest side which in deed is the perpendicular distance.

$\Rightarrow (y-\hat{y})$ is orthogonal to the column space of X. Hence proved.