06 / 05 / 2022

Rui Chen

Homework # 1

1. When the person has height of 150, the nearest neighbors are 150, 168, and 171. Then the estimate weight is $\hat{y}_{KNN} = (65 + 78 + 80)/3 = 74.33$

When the person has height of 155, the nearest neighbors are 150, 168, and 171. Then the estimate weight is $\hat{y}_{KNN} = (65 + 78 + 80)/3 = 74.33$

When the person has height of 165, the nearest neighbors are 168, 171, and 178. Then the estimate weight is $\hat{y}_{KNN} = (78 + 80 + 83)/3 = 80.33$

When the person has height of 190, the nearest neighbors are 191, 182, and 178. Then the estimate weight is $\hat{y}_{KNN} = (100 + 80 + 83)/3 = 87.67$

2. When the person has height of 150, the first 3 nearest neighbors are 150, 168, and 171. Since distance between 150 and 150 is 0, the $\hat{y}_{KNN} = 65$.

When the person has height of 155, the first 3 nearest neighbors are 150, 168, and 171 with weight $\frac{1}{5}$, $\frac{1}{13}$, and $\frac{1}{16}$. Then $\hat{y}_{KNN} = (\frac{1}{5} \times 65 + \frac{1}{13} \times 78 + \frac{1}{16} \times 80) / (\frac{1}{5} + \frac{1}{13} + \frac{1}{16}) = 70.71$

When the person has height of 165, the first 3 nearest neighbors are 168, 171 and 178 with weight $\frac{1}{3}$, $\frac{1}{6}$, and $\frac{1}{13}$. Then $\hat{y}_{KNN} = (\frac{1}{3} \times 78 + \frac{1}{6} \times 80 + \frac{1}{13} \times 83) / (\frac{1}{3} + \frac{1}{6} + \frac{1}{13}) = 79.24$

When the person has height of 190, the first 3 nearest neighbors are 191, 182, and 178. with weight 1, $\frac{1}{8}$, and $\frac{1}{12}$. Then $\hat{y}_{KNN} = (1 \times 100 + \frac{1}{8} \times 80 + \frac{1}{12} \times 83) / (1 + \frac{1}{8} + \frac{1}{12}) = 96.76$.

3.
$$\nabla_x J(x) = \begin{bmatrix} \frac{dJ(x)}{dx_1} \\ \frac{dJ(x)}{dx_2} \\ \vdots \\ \frac{dJ(x)}{dx_n} \end{bmatrix} = \begin{bmatrix} \frac{d(x^TQx + d^Tx + c)}{dx_1} \\ \frac{d(x^TQx + d^Tx + c)}{dx_2} \\ \vdots \\ \frac{d(x^TQx + d^Tx + c)}{dx_n} \end{bmatrix} = \begin{bmatrix} \frac{d(x^TQx)}{dx_1} + \frac{d(d^Tx)}{dx_1} + \frac{dc}{dx_1} \\ \vdots \\ \frac{d(x^TQx)}{dx_n} + \frac{d(d^Tx)}{dx_n} + \frac{dc}{dx_n} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{d(x^TQx)}{dx_1} \\ \frac{d(x^TQx)}{dx_2} \\ \vdots \\ \frac{d(x^TQx)}{dx_n} \end{bmatrix} + \begin{bmatrix} \frac{d(d^Tx)}{x_1} \\ \frac{d(d^Tx)}{x_2} \\ \vdots \\ \frac{d(d^Tx)}{x_n} \end{bmatrix} + 0 = 2Q^Tx + d$$

since $Q = Q^T$, then

$$\nabla_x J(x) = 2Qx + d.$$

By difinition, $H_{ij} = \frac{d^2J}{dx_i dx_j}$, then the Hessian matrix of J is

$$H = \begin{bmatrix} \frac{d^2J}{dx_1^2} & \frac{d^2J}{dx_1 dx_2} & \cdots & \frac{d^2J}{dx_1 dx_n} \\ \vdots & \ddots & & \\ & & & \\ \frac{d^2J}{dx_n dx_1} & & \cdots & \frac{d^2J}{dx_n^2} \end{bmatrix} = \frac{d\, \nabla_x J(x)}{dx} = 2Q.$$

4. $\hat{y} = X(X^TX)^{-1}X^Ty$

For linear regression model, the main idea is find the ~~linear~~ line which can ~~min~~ minimun the distances between each points and ~~line~~ line. KNN is a method that find the first k nearest neighbours and estimate the test instance. Therefore, if the k is appropriate, the prediction of linear regression model is a special case of KNN.

5. Suppose the system is $y = Xa$ where $y \in \mathbb{R}^n$, $X \in \mathbb{R}^{n\times(p+1)}$, and $a \in \mathbb{R}^{(p+1)}$

~~Then $\hat{a} = (A^TA)^{-1}A^Tb$ and $\hat{a}$ is size of~~

Then $\hat{a} = (X^TX)^{-1}X^Ty$ and $\hat{a}$ is size of $(p+1)\times 1$.

Therefore $\hat{y} = X\hat{a} = X(X^TX)^{-1}X^Ty$ and it is a column ~~of~~ space of X.

6. $\hat{y} = X(X^TX)^{-1}X^Ty$.

Suppose the column space of X is represented as Xa.

$(y-\hat{y})\cdot Xa = yXa - X(X^TX)^{-1}X^TyXa = yXa - yXa = 0$.

Therefore, $y-\hat{y}$ is orthogonal to the column space of X.