Prob. 1

Monday, June 6, 2022 4:38 PM

1. Assume that you are given the following sample . Estimate the weight of people whose heights are 150, 155, 165, and 190 cm, using KNN with $k = 3$:

$$\hat{y}_{KNN} = \frac{y_1 + y_2 + \cdots + y_k}{k}$$

where $y_1, y_2, \cdots, y_k$ are the labels of the k nearest neighbors to your test instance. (10 pts)

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

① Calc a person's weight w/ height 150

Euclidean distance = d = $\sqrt{(h_t - h_{obs})^2}$

$d_1 = \sqrt{(150-171)^2} = \sqrt{(-21)^2} = 21$

$d_2 = \sqrt{(150-168)^2} = 18$

$d_3 = \sqrt{(150-191)^2} = 41$

$d_4 = \sqrt{(150-182)^2} = 32$

$d_5 = \sqrt{(150-150)^2} = 0$

$d_6 = \sqrt{(150-178)^2} = 28$

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	**171**	**21**	**80**
2	**168**	**18**	**78**
3	191	41	100
4	182	32	80
5	**150**	**0**	**65**
6	178	28	83
7	150	N/A	74.33

② find 3 near neighbrs as k = 3, so from the table, people with heights 171, 168, 150

Therefore, to calc the est weight $\hat{y}_{KNN}$ is

$$\hat{y}_{KNN} = \frac{y_1 + y_2 + y_3}{3} = \frac{80 + 78 + 65}{3} = 74.33$$

⊛ repeat the same calculation

⊛ Calc a person's weight w/ height 155

Euclidean distance = $d = \sqrt{(h_t - h_{obs})^2}$

$d_1 = \sqrt{(155 - 171)^2} = 16$

$d_2 = \sqrt{(155 - 168)^2} = 13$

$d_3 = \sqrt{(155 - 191)^2} = 36$

$d_4 = \sqrt{(155 - 182)^2} = 27$

$d_5 = \sqrt{(155 - 150)^2} = 5$

$d_6 = \sqrt{(155 - 178)^2} = 23$

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	**171**	**16**	**80**
2	**168**	**13**	**78**
3	191	36	100
4	182	27	80
5	**150**	**5**	**65**
6	178	23	83
7	155	N/A	74.33

⊛ find 3 near neighbrs as $k = 3$, so from the table, people with heights 171, 168, 150

Therefore, to calc the est weight $\hat{y}_{KNN}$ is

$$\hat{y}_{KNN} = \frac{y_1 + y_2 + y_3}{3} = \frac{80 + 78 + 65}{3} = 74.33$$

⊛ repeat the same calculation

(*) calc a person's weight w/ height 165

Euclidean distance = d = $\sqrt{(h_t - h_{obs})^2}$

$d_1 = \sqrt{(165-171)^2} = 6$

$d_2 = \sqrt{(165-168)^2} = 3$

$d_3 = \sqrt{(165-191)^2} = 26$

$d_4 = \sqrt{(165-182)^2} = 17$

$d_5 = \sqrt{(165-150)^2} = 15$

$d_6 = \sqrt{(165-178)^2} = 13$

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	**171**	**6**	**80**
2	**168**	**3**	**78**
3	191	26	100
4	182	17	80
5	150	15	65
6	**178**	**13**	**83**
7	165	N/A	80.33

(*) find 3 near neighbrs as k = 3, so from the table, people with heights 171, 168, 178

Therefore, to calc the est weight $\hat{y}_{kNN}$ is

$$\hat{y}_{kNN} = \frac{y_1 + y_2 + y_3}{3} = \frac{80 + 78 + 83}{3} = 80.33$$ ✓✓

(*) repeat the same calculation

(*) calc a person's weight w/ height 165

Euclidean distance = d = $\sqrt{(h_t - h_{obs})^2}$

$d_1 = \sqrt{(190-171)^2} = 19$

$d_2 = \sqrt{(190-168)^2} = 22$

$d_3 = \sqrt{(190-191)^2} = 1$

$$d_3 = \sqrt{(190 - 191)^2} = 1$$

$$d_4 = \sqrt{(190 - 182)^2} = 8$$

$$d_5 = \sqrt{(190 - 150)^2} = 40$$

$$d_6 = \sqrt{(190 - 178)^2} = 12$$

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	171	19	80
2	168	22	78
3	**191**	**1**	**100**
4	**182**	**8**	**80**
5	150	40	65
6	**178**	**12**	**83**
7	190	N/A	87.67

③ find 3 near neighbrs as $k=3$, so from the table, people with heights 191, 182, 178

Therefore, to calc the est weight $\hat{y}_{KNN}$ is

$$\hat{y}_{KNN} = \frac{y_1 + y_2 + y_3}{3} = \frac{100 + 80 + 83}{3} = 87.67$$

Prob. 2

Monday, June 6, 2022 7:35 PM

2. Repeat 1, but instead of using the simple average of the labels of k nearest neighbors, which is use the following weighted average:

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

where the weight w_i for the label y_i of instance i is determined as $1/d_i$, where d_i the distance between the instance i and the test instance. (10 pts)

(*) the distances for Person's weight 150

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	**171**	**21**	**80**
2	**168**	**18**	**78**
3	191	41	100
4	182	32	80
5	**150**	**0**	**65**
6	178	28	83
7	150	N/A	66.3

$w_1 = \frac{1}{d_1} = \frac{1}{21}$, $w_2 = \frac{1}{d_2} = \frac{1}{18}$, $w_3 = \frac{1}{d_3} = \frac{1}{0} = \infty$

Note: w_3 is a large number, and since $w_i \in (0,1)$, we can assume that $w_3 = 1$

Now, calculate the estimated weight

$$\hat{y}_{KNN} = \frac{w_1 y_1 + w_2 y_2 + w_3 y_3}{w_1 + w_2 + w_3} = \frac{80/21 + 78/18 + 65}{\frac{1}{21} + \frac{1}{18} + 1} = 66.3$$

(*) the distances for Person's weight 155

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	**171**	**16**	**80**
2	**168**	**13**	**78**
3	191	36	100
4	182	27	80
5	**150**	**5**	**65**
6	178	23	83
7	155	N/A	66.3

$w_1 = \frac{1}{21}$, $w_2 = \frac{1}{18}$, $w_3 = 1$ → by assumption

Now, calculate the estimated weight

$$\hat{y}_{kNN} = \frac{w_1 y_1 + w_2 y_2 + w_3 y_3}{w_1 + w_2 + w_3} = \frac{80/21 + 78/18 + 65}{\frac{1}{21} + \frac{1}{18} + 1} = 66.3 \checkmark\checkmark$$

(*) the distances for Person's weight 165

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	**171**	**6**	**80**
2	**168**	**3**	**78**
3	191	26	100
4	182	17	80
5	150	15	65
6	**178**	**13**	**83**
7	165	N/A	79.24

$w_1 = \frac{1}{6}$, $w_2 = \frac{1}{3}$, $w_3 = \frac{1}{13}$

Now, calculate the estimated weight

$$\hat{y}_{kNN} = \frac{w_1 y_1 + w_2 y_2 + w_3 y_3}{w_1 + w_2 + w_3} = \frac{80/6 + 78/3 + 83/13}{\frac{1}{6} + \frac{1}{3} + \frac{1}{13}} = 79.24$$

(*) the distances for Person's weight 190

Person	Height (cm)	Distance (Euclidean)	Weight (kg)
1	171	19	80
2	168	22	78
3	**191**	**1**	**100**
4	**182**	**8**	**80**
5	150	40	65
6	**178**	**12**	**83**
7	190	N/A	96.76

$w_1 = 1$, $w_2 = 1/8$, $w_3 = 1/12$

Now, calculate the estimated weight

$$\hat{y}_{kNN} = \frac{w_1 y_1 + w_2 y_2 + w_3 y_3}{w_1 + w_2 + w_3} = \frac{100 + 10 + 83/12}{1 + \frac{1}{8} + \frac{1}{12}} = 96.76$$

Prob. 3

Monday, June 6, 2022 7:35 PM

3. Assume that $J(\mathbf{x}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{d}^T\mathbf{x} + c$ where $\mathbf{Q} = \mathbf{Q}^T \in \mathbb{R}^{n\times n}$, $\mathbf{x}, \mathbf{d} \in \mathbb{R}^n$, and $c \in \mathbb{R}|$. Show that $\nabla_{\mathbf{x}} J(\mathbf{x}) = 2\mathbf{Q}\mathbf{x} + \mathbf{d}$ and $\mathbf{H} = \frac{\partial^2 J}{\partial \mathbf{x} \partial \mathbf{x}^T} = 2\mathbf{Q}$. $\mathbf{H}_{ij} = \frac{\partial^2 J}{\partial x_i \partial x_j}$ and $\mathbf{H}$ is called the Hessian matrix of J. (10 pts)

Prove $\nabla_x J(x) = 2Qx$,

$$\nabla_x J(x) = \frac{\partial J}{\partial x} = \frac{\partial}{\partial x}\left(x^T Q x + d^T x + c\right)$$

since Q is symmetric and the function is linear, we can treat $x^T Q x$ as $q x^2$ and $d^T x = dx$

$$\therefore \nabla_x J(x) = \frac{\partial J}{\partial x} = 2Qx + d$$

prove $H = \frac{\partial^2 J}{\cancel{\partial x \partial x'}} = 2Q$

$$\frac{\partial^2 J}{\partial x \partial x^T} = \frac{\partial}{\partial x^T}\left(2Qx + d\right) = \frac{\partial}{\partial x^T}\left[\left(2x^T Q^T + d^T\right)\right]$$

$$= \frac{\partial}{\partial x^T}\left[\left(2x^T Q + d^T\right)\right] = 2Q$$

Prob. 4

Monday, June 6, 2022 7:37 PM

4. Write down the prediction $\widehat{\mathbf{y}}$ for a test row vector $\mathbf{x}'_{1\times p}$ made by a linear regression model in terms of $\mathbf{y}$ the vector of labels of the training set and $\mathbf{X}_{n\times(p+1)}$, the (augmented) feature matrix, and explain why $\widehat{\mathbf{y}}$ can be viewed as a special case of KNN regression. (10 pts)

$$\hat{y} = X\hat{\beta} = \underbrace{X(X^TX)^{-1}X^T}_{H}y = Hy$$

The linear regression model is a form of KNN. We can define KNN Regression model as $y_{KNN} = \frac{y}{k}$ where $y = [y_1, y_2, \dots, y_k]$. Comparing to the linear regression model $\hat{y}$, simply, $H = \frac{1}{k}$ if x_i's are the nearest $\underline{k}$ point and zero elsewhere.

let's take $y_0 = \hat{f}(x_0) = [x_0 \;\; 1]\hat{\beta} = \underbrace{[x_0 \;\; 1](X^TX)^{-1}X^T}_{H}y$

we can define H as

$$H = (x_0 \;\; 1)\,(X^TX)^{-1}\begin{pmatrix} 1 \\ x_i \end{pmatrix}, \quad i \in (1, 2, \dots n) \longrightarrow ☆$$

Simplify X^TX

$$X^TX = \begin{pmatrix} \cdots 1 \cdots \\ \cdots x_i \cdots \end{pmatrix}_{2\times n}\begin{pmatrix} \vdots & \vdots \\ 1 & x_i \\ \vdots & \vdots \end{pmatrix}_{n\times 2} = \begin{pmatrix} \sum 1 & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

To find the inverse, $|X^TX|$

$$|X^T X| = n \sum x_i^2 - \sum x_i^2 = (n-1) \sum x_i^2$$

$$\therefore (X^T X)^{-1} = \frac{1}{(n-1)\sum x_i^2} \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix} \rightarrow ①$$

plug ① into ☆ :

$$\underbrace{H = \frac{1}{(n-1)\sum x_i^2} (x_0 \quad 1) \begin{pmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{pmatrix}_{2\times 2} \begin{pmatrix} -1- \\ -x_i- \end{pmatrix}_{2\times n}}$$

This should equal $\frac{1}{n}$ but it seems I missed some steps

$$= \frac{1}{(n-1)\sum x_i^2} (x_0 \quad 1) \begin{pmatrix} \sum x_i^2 - \sum x_i^2 \\ -\sum x_i + n \sum x_i \end{pmatrix}$$

$$= \frac{1}{\cancel{(n-1)} \sum x_i^2} (x_0 \quad 1) \begin{pmatrix} 0 \\ \cancel{(n-1)} \cancel{\sum x_i} \end{pmatrix} = \frac{1}{\sum x_i} (x_0 \quad 1) \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$= \frac{1}{\sum_{i=1}^{n} x_i}$$

Prob. 5

Monday, June 6, 2022 7:37 PM

5. Show that the for $\mathbf{y} \in \mathbb{R}^n$, $\hat{\mathbf{y}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ is a member of the column space of $\mathbf{X}$, i.e. is a linear combination of the columns of $\mathbf{X} \in \mathbb{R}^{n\times(p+1)}$. (10 pts)

$X \in R^{n\times(p+1)}$, where $(p+1) \leq n$

$$H_{n\times n} = \underset{n\times(p+1)}{X} \underset{(p+1)\times(p+1)}{(X^T X)^{-1}} X^T_{(p+1)\times n}$$

note that $H_{n\times n}$ has a rank of $p+1$

suppose we have two vectors:

i) v is the column space of $\underline{X}$

ii) w is orthogonal to $Col(X)$

then we can prove from (i) and (ii) the following

i) suppose $\underset{n\times 1}{v} = \underset{n\times(p+1)}{X} \underset{(p+1)\times 1}{u}$

$$\therefore Hv = X(X^T X)^{-1} X^T v = X(X^T X)^{-1} X^T X u$$

$$= X\,(X^{-1}(X^T)^{-1})\,X^T X u = (\underbrace{X X^{-1}}_{I})\,(\underbrace{X^{-1} X}_{I})^T X u = Xu = v$$

ii) $\underset{n\times n}{H}\underset{n\times 1}{w} = X(X^T X)^{-1} \underbrace{X^T w}_{0} = X(X^T X)^{-1}\, 0 = 0$

So we conclude that

$Hv = \lambda_1 v$ where $\lambda_1 = 1$? ... ...

$H v = \lambda_1 v$ where $\left.\begin{array}{l}\lambda_1 = 1 \\ \lambda_2 = 0\end{array}\right\}$ e-values

$H w = \lambda_2 w$

and v, w are e-vectors, which span the whole space of $X \in R^{n \times p+1}$

Prob. 6

Monday, June 6, 2022 7:37 PM

6. Show that in linear regression, if $\widehat{\beta}$ minimizes RSS(β), then $\mathbf{y} - \widehat{\mathbf{y}}$ is orthogonal to the column space of $\mathbf{X}$. (10 pts)

$$y = X\beta + \varepsilon = \begin{bmatrix} 1 & x_{11} & \cdots & x_{np} \\ 1 & x_{21} & & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

In order to minimize RSS(β)

$$RSS(\beta) = \varepsilon^T\varepsilon = (y - X\beta)^T(y - X\beta)$$

$$= (y^T - \beta^T X^T)(y - X\beta)$$

$$= y^T y \underbrace{- y^T X\beta - \beta^T X^T y}_{-2y^T X\beta} + \beta^T X^T X\beta$$

$$= y^T y - 2y^T X\beta + \beta^T X^T X\beta$$

$$\therefore \frac{\partial RSS(\beta)}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\cancel{y^T y} - 2y^T X\beta + \beta^T X^T X\beta\right)$$

$$= -2y^T X + 2\beta^T X^T X$$

$$-2y^T X + 2\beta^T X^T X = 0$$

$$\cancel{2}\beta^T X^T X = \cancel{2} y^T X$$

take transpose on both sides $\rightarrow$ $X^T X\beta = X^T y$ $\rightarrow$ Normal equation assume $X^T X$ is linearly independent, then

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

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

$\therefore X \varepsilon = X(y - \hat{y}) = X(y - X\hat{\beta}) = 0 \rightarrow$ orthogonality