Assignment 7 Solution

Nayan Man Singh Pradhan

Excercise 1. (Training error)

In this task, we consider the training data

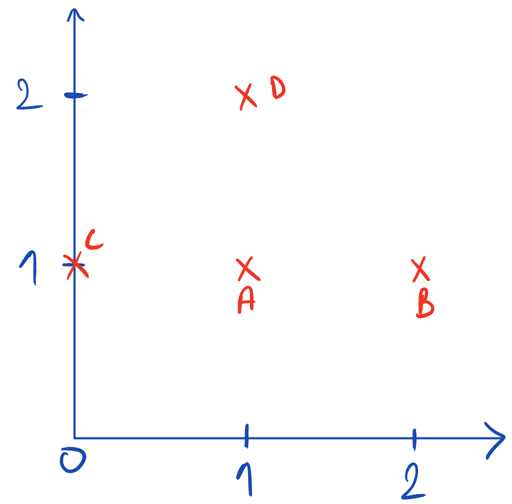
$$\mathcal{T} = \{((1,1)^\top, 2), ((2,1)^\top, 5), ((0,1)^\top, 1), ((1,2)^\top, 5)\}.$$

a) Build a predictor using kNN regression for $k = 3$ and evaluate the training error of the predictor.

b) Build a predictor using linear regression by least squares and evaluate the training error of the linear model.

(4 Points)

a) let $A(1,1)$, $B(2,1)$, $C(0,1)$ and $D(1,2)$



Computing the distance between the points

	A	B	C	D
A	0	1	1	1
B	1	0	2	1.414
C	1	2	0	1.414
D	1	1.414	1.414	0

Using KNN regression (with $k=3$)

For $(1,1)$, the closest 3 neighbors are $(1,1)$, $(2,1)$, $(0,1)$

$$Y(1,1) = \frac{1}{3}(2+5+1) = \frac{8}{3} = 2.67$$

For $(2,1)$, the closest 3 neighbors are $(2,1)$, $(1,1)$, $(1,2)$

$$Y(2,1) = \frac{1}{3}(5+2+5) = \frac{12}{3} = 4$$

For $(0,1)$, the closest 3 neighbors are $(0,1)$, $(1,1)$, $(1,2)$

$$Y(0,1) = \frac{1}{3}(1+2+5) = \frac{8}{3} = 2.67$$

For $(1,2)$, the closest 3 neighbors are $(1,2)$, $(1,1)$, $(2,1)$

$$Y(1,2) = \frac{1}{3}(5+2+5) = \frac{12}{3} = 4$$

Training Error:

$$TE(f,T) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i))$$

$$= \frac{1}{4}\sum_{i=1}^{4} L_2(y_i, f(x_i))$$

$$= \frac{1}{4}\left((2-2.67)^2+(5-4)^2+(1-2.67)^2+(5-4)^2\right)$$

$$= 1.309$$

b) $X = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}$, $X^T = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}$, $y = \begin{pmatrix} 2 \\ 5 \\ 1 \\ 5 \end{pmatrix}$

$$X^TX = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}_{3\times4} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{pmatrix}_{4\times3} = \begin{pmatrix} 4 & 4 & 5 \\ 4 & 6 & 5 \\ 5 & 5 & 7 \end{pmatrix}$$

$$X^Ty = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}_{3\times4} \cdot \begin{pmatrix} 2 \\ 5 \\ 1 \\ 5 \end{pmatrix}_{4\times1} = \begin{pmatrix} 13 \\ 17 \\ 18 \end{pmatrix}$$

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

$$\begin{pmatrix} 4 & 4 & 5 \\ 4 & 6 & 5 \\ 5 & 5 & 7 \end{pmatrix} \hat{\beta} = \begin{pmatrix} 13 \\ 17 \\ 18 \end{pmatrix}$$

Using gaussian elimination to solve,

$$\left(\begin{array}{ccc|c} 4 & 4 & 5 & 13 \\ 4 & 6 & 5 & 17 \\ 5 & 5 & 7 & 18 \end{array}\right)$$

$$\left(\begin{array}{ccc|c} 4 & 4 & 5 & 13 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 3/4 & 7/4 \end{array}\right)$$

$$\frac{3}{4}\beta_3 = \frac{7}{4} \Rightarrow \beta_3 = \frac{7}{3} \,//$$

$$2\beta_2 = 4 \Rightarrow \beta_2 = 2 \,//$$

$$4\beta_1 + 4\beta_2 + 5 \times \beta_3 = 13 \Rightarrow 4\beta_1 + 8 + \frac{35}{3} = 13 \Rightarrow \beta_1 = -\frac{5}{3} \,//$$

$$\therefore f(\beta) = -\frac{5}{3} + 2\beta_1 + \frac{7}{3}\beta_2$$

$$f(1,1) = -\frac{5}{3} + 2 + \frac{7}{3} = 2.67$$

$$f(2,1) = -\frac{5}{3} + 2 \times 2 + \frac{7}{3} = 4.67$$

$$f(0,1) = -\frac{5}{3} + 0 + \frac{7}{3} = 0.67$$

$$f(1,2) = -\frac{5}{3} + 2 \times 1 + \frac{7}{3} \times 2 = 5$$

Training Error:

$$TE(f, T) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i))$$

$$= \frac{1}{4}\sum_{i=1}^{4} L_2(y_i, f(x_i))$$

$$= \frac{1}{4}\left((2-2.67)^2 + (5-4.67)^2 + (1-0.67)^2 + (5-5)^2\right)$$

$$= 0.167 \,//$$

Excercise 2. (Cross validation)

In this task, we consider the data set

$$\mathcal{T} = \{(-2,4), (2,4), (1,1), (-1,1), (0,0), (3,9)\} .$$

a) Evaluate the (expected) generalization error of the kNN regressor with $k = 2$ by $\mathcal{K}$-fold cross validation with $\mathcal{K} = 3$. (Do a deterministic, i.e. not randomized, splitting of the given data following the ordering of the samples.

b) Evaluate the (expected) generalization error of the linear model by leave-one-out cross validation. (It is fine to use the help of a computer to fit the individual linear models.)

(4 Points)

a) $T = \{(-2,4), (2,4), (1,1), (-1,1), (0,0), (3,9)\}$

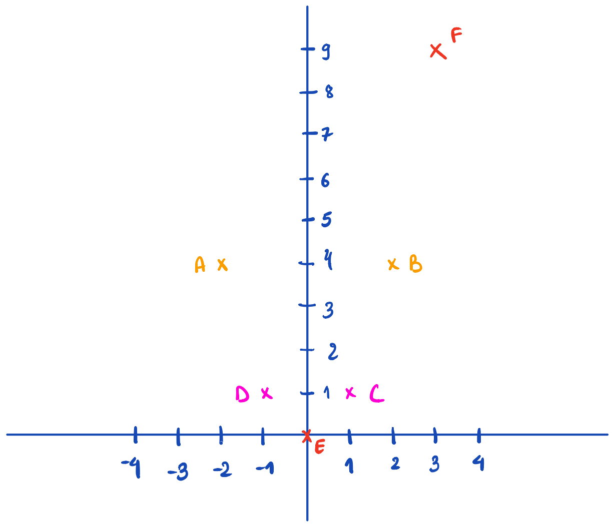
$K = 3$

Let,

$T_1 = \{(-2,4), (2,4)\}$ $\quad T_1 = A(-2,4), B=(2,4)$

$T_2 = \{(1,1), (-1,1)\}$ $\quad T_2 = C(1,1), D=(-1,1)$

$T_3 = \{(0,0), (3,9)\}$ $\quad T_3 = E(0,0), F(3,9)$



(i) Using T_1 and T_2 as training and T_3 as cross validation set

For E, the 2 nearest neighbors are: C and D

$= \frac{1}{2} \times (1+1) = 1$

For F, the 2 nearest neighbors are B and A

$= \frac{1}{2} \times (4+4) = 4$

$\text{Error}_1 = \frac{1}{2} \times \left[(4-9)^2 + (1-0)^2 \right] = 13$

(ii) Using T_2 and T_3 as training and T_1 as cross validation set
for B, the two nearest neighbors are C and D.

$$= \frac{1}{2} \times (1+1) = 1$$

for A, the two nearest neighbors are D and E

$$= \frac{1}{2} \times (1+0) = 0.5$$

$$\text{Error}_2 = \frac{1}{2} \times \left[(1-4)^2 + (0.5-4)^2\right] = \frac{21.25}{2} = 10.625$$

(iii) Using T_1 and T_3 as training and T_2 as cross-validation set
for C, the two nearest neighbors are E and B

$$= \frac{1}{2} \times (4+0) = 2$$

for D, the two nearest neighbors are A and E

$$= \frac{1}{2} \times (4+0) = 2$$

$$\text{Error}_3 = \frac{1}{2} \times \left[(2-1)^2 + (1-2)^2\right] = \frac{2}{2} = 1$$

$$\text{Expected generalization error} = \frac{1}{3} \times (13 + 10.625 + 1) = 8.208$$

b) $T = \{(-2,4), (2,4), (1,1), (-1,1), (0,0), (3,9)\}$

let, $T_1(-2,4)$, $T_2(2,4)$, $T_3(1,1)$, $T_4(-1,1)$, $T_5(0,0)$, $T_6(3,9)$

① Consider T_1, T_2, T_3, T_4, T_5 as training and T_6 as cross-validation set.

$$X = \begin{pmatrix} 1 & -2 \\ 1 & 2 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}, \quad X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & 2 & 1 & -1 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 4 \\ 4 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & 2 & 1 & -1 & 0 \end{pmatrix}_{2x5} \begin{pmatrix} 1 & -2 \\ 1 & 2 \\ 1 & 1 \\ 1 & -1 \\ 1 & 0 \end{pmatrix}_{5x2} = \begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix}$$

$$X^T y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ -2 & 2 & 1 & -1 & 0 \end{pmatrix}_{2x5} \begin{pmatrix} 4 \\ 4 \\ 1 \\ 1 \\ 0 \end{pmatrix}_{5x1} = \begin{pmatrix} 10 \\ 0 \end{pmatrix}_{2x1}$$

$$X^T X \beta = X^T y$$

$$\begin{pmatrix} 5 & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 0 \end{pmatrix}$$

$5\beta_1 + 0 = 10$ $\qquad$ $0\beta_1 + 10\beta_2 = 0$

$\beta_1 = 2$ $\qquad$ $\beta_2 = 0$

$Y_p = 2 \cdot + 0$
$= 2$

Testing $T_6(3, 9)$

$Y_{pred} = 2$
$Y_{Real} = 9$

$Error_1 = (y_R - y_P)^2 = (9-2)^2 = 49$

*The other computations are done using the implemented algorithm from previous programming assignments.

(ii) Consider $T_1, T_2, T_3, T_4,$ and T_6 as training and T_5 as cross-validation set.

$T_5\ (0,0)$

$Y_{pred} = 3.26$

$Y_{Real} = 0$

$$Error_2 = (Y_{Real} - Y_{pred})^2 = (0-3.26)^2 = 10.627$$

(iii) Consider $T_1, T_2, T_3, T_5,$ and T_6 as training and T_4 as cross-validation set.

$T_4\ (-1, 1)$

$Y_{pred} = 1.94$

$Y_{Real} = 1$

$$Error_3 = (Y_{Real} - Y_{pred})^2 = (1-1.94)^2 = 0.8836$$

(iv) Consider $T_1, T_2, T_4, T_5,$ and T_6 as training and T_3 as cross-validation set.

$T_3\ (1, 1)$

$Y_{pred} = 4.26$

$Y_{Real} = 1$

$$Error_4 = (Y_{Real} - Y_{pred})^2 = (1-4.26)^2 = 10.627$$

(v) Consider $T_1, T_3, T_4, T_5,$ and T_6 as training and T_2 as cross-validation set.

$T_2\ (2, 4)$

$Y_{pred} = 4.94$

$Y_{real} = 4$

$$Error_5 = (Y_{Real} - Y_{pred}) = (4-4.94)^2 = 0.8836$$

(vi) Consider $T_2, T_3, T_4, T_5,$ and T_6 as training and T_1 as cross-validation set.

$T_1\ (-2, 4)$

$Y_{pred} = -3$

$Y_{real} = 4$

$$Error = (Y_{Real} - Y_{pred}) = (4+3)^2 = 49$$

$$\text{Expected generalization error} = \frac{1}{6} \times (49 + 10.627 + 0.8836 + 10.627 + 0.8836 + 49)$$

$$= 20.17$$

Excercise 3. (Observations on the training error)

Provide quantitative reasoning to the questions below.

a) What is the training error of kNN regression with neighbourhood size $k = 1$? Give the result and an explanation of how you get to the result.

b) What is the training error of linear regression with an output dimension of $K = 1$ for $N = 2$ distinct training samples? Give the result and an explanation of how you get to the result.

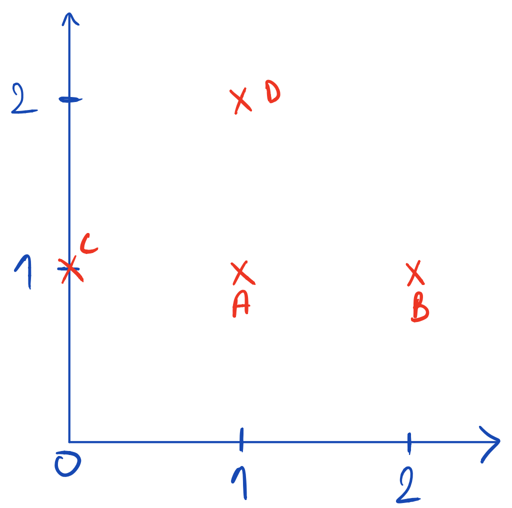
(4 Points)

a) The training error of KNN regression with neighbourhood size $K=1$ is 0. The training error is calculated as the average of the loss function in the training set given that we train our regressor using all training samples. If we test our KNN regression algorithm with respect to only $k=1$ nearest neighbour, the algorithm is always going to pick the exact same sample. Because the regressor has been trained using the sample and other neighbours (beside the sample) are not considered, there would be no training error.

Eg: Consider the training data set:

$$T: \{ ((1,1)^T, 2), ((2,1)^T, 5), ((0,1)^T, 1), ((1,2)^T, 5) \}$$

let $A(1,1)$, $B(2,1)$, $C(0,1)$ and $D(1,2)$



Computing the distance between the points

	A	B	C	D
A	0	1	1	1
B	1	0	2	1.414
C	1	2	0	1.414
D	1	1.414	1.414	0

Using KNN regression (with $k = 1$)

For $(1,1)$, the closest neighbor is $(1,1)$

$$Y(1,1) = \frac{1}{1}(2) = 2$$

For $(2,1)$, the closest neighbor is $(2,1)$

$$Y(2,1) = \frac{1}{1}(5) = 5$$

for $(0,1)$, the closest neighbor is $(0,1)$

$$Y(0,1) = \frac{1}{1}(1) = 1$$

For $(1,2)$, the closest neighbor is $(1,2)$

$$Y(1,2) = \frac{1}{1}(5) = 5$$

Training Error:

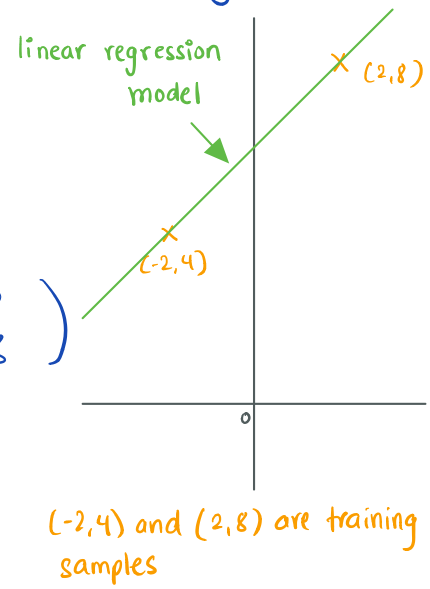
$$TE(f,T) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, f(x_i))$$

$$= \frac{1}{4}\left((\cancelto{0}{2-2})^2 + (\cancelto{0}{5-5})^2 + (\cancelto{0}{1-1})^2 + (\cancelto{0}{5-5})^2 \right)$$

$$= 0.$$

Hence we can conclude that the training error of KNN regression with neighbourhood size $k=1$ is 0.

b) The training error of linear regression with an output dimension of $K=1$ for $N=2$ training samples will be 0. When we take into account only $N=2$ training samples, the linear regression algorithm simply connects the two points with a linear line that can be represented with the equation $Y=\beta_1+\beta_2 X$. This is an output dimension of $K=1$. Later, when we test the two points that we used for training the model, the predicted Y value lies on the linear line as the line is simply created by joining the two points.



Eg: Consider training samples:

$$T=\{(-2,4),(2,8)\}$$

$$X=\begin{pmatrix}1 & -2\\ 1 & 2\end{pmatrix},\quad X'=\begin{pmatrix}1 & 1\\ -2 & 2\end{pmatrix},\quad y=\begin{pmatrix}4\\ 8\end{pmatrix}$$

$$X^TX=\begin{pmatrix}1 & 1\\ -2 & 2\end{pmatrix}\begin{pmatrix}1 & -2\\ 1 & 2\end{pmatrix}=\begin{pmatrix}2 & 0\\ 0 & 8\end{pmatrix}$$

$$X^Ty=\begin{pmatrix}1 & 1\\ -2 & 2\end{pmatrix}\begin{pmatrix}4\\ 8\end{pmatrix}=\begin{pmatrix}12\\ 8\end{pmatrix}$$

$$X^TX\hat{\beta}=X^Ty$$

$$\begin{pmatrix}2 & 0\\ 0 & 8\end{pmatrix}\begin{pmatrix}\hat{\beta}_1\\ \hat{\beta}_2\end{pmatrix}=\begin{pmatrix}12\\ 8\end{pmatrix}$$

$$\hat{\beta}_1=6 \qquad \hat{\beta}_2=1$$

$$\therefore Y=\hat{\beta}_1+X\hat{\beta}_2$$

$$Y_{pred}=6+X$$

For point $(-2,4)$, $Y_{pred}=6-2=4$

For point $(2,8)$, $Y_{pred}=6+2=8$

$\therefore$ Training Error $=\frac{1}{2}\left((4-4)^2+(8-8)^2\right)=0$