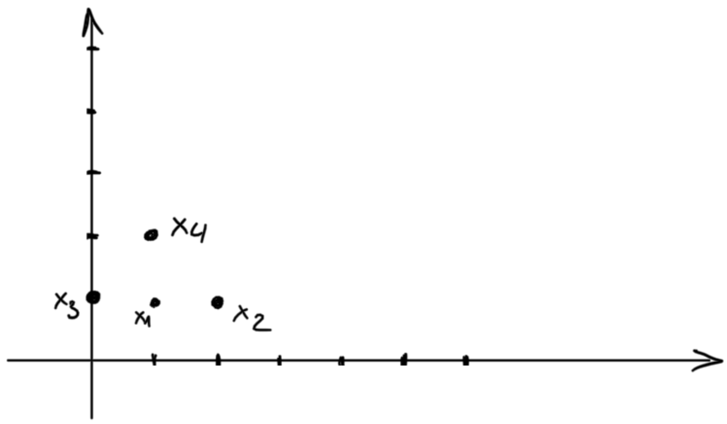
Exercise 1 | Assignment #7

$$\mathcal{T} = \left\{ \left((1,1)^t, 2\right), \left((2,1), 5\right), \left((0,1)^t, 1\right), \left((1,2)^t, 5\right) \right\}$$

a|



$$f_{kNN}(x_1) = \frac{y(x_1) + y(x_2) + y(x_3)}{3} = \frac{2+5+1}{3} = \frac{8}{3}$$

$$f_{kNN}(x_2) = \frac{y(x_1) + y(x_2) + y(x_4)}{3} = \frac{2+5+5}{3} = 4$$

$$f_{kNN}(x_3) = \frac{y(x_1) + y(x_4) + y(x_3)}{3} = \frac{2+5+1}{3} = \frac{8}{3}$$

$$f_{kNN}(x_4) = \frac{y(x_4) + y(x_1) + y(x_3)}{3} = \frac{5+2+1}{3} = \frac{8}{3}$$

$$TE_{kNN} = \frac{1}{|\mathcal{T}|} \sum_{i=1}^{|\mathcal{T}|} \left(f_{kNN}(x_i) - y(x_i) \right)^2 = \frac{1}{4} \left(\left(\frac{8}{3} - 2\right)^2 + (4-5)^2 + \left(\frac{8}{3} - 1\right)^2 + \left(\frac{8}{3} - 5\right)^2 \right)$$

$$= \frac{1}{4} \left(\frac{4}{9} + 1 + \frac{25}{9} + \frac{49}{9} \right) = \frac{87}{36} = \boxed{\frac{29}{12}}$$

b|

$$X^T X \beta = X^T Y$$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 2 \end{bmatrix} \vec{\beta} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 5 \\ 1 \\ 5 \end{bmatrix} \Rightarrow \left[\begin{array}{ccc|c} 4 & 4 & 5 & 13 \\ 4 & 6 & 5 & 17 \\ 5 & 5 & 7 & 18 \end{array}\right] \Rightarrow$$

$$\left[\begin{array}{ccc|c} 4 & 4 & 5 & 13 \\ 0 & 2 & 0 & 4 \\ 0 & 0 & 3/4 & 7/4 \end{array}\right] \xRightarrow{\text{Back Substitution}} \beta_2 = \frac{7}{3} \quad \beta_1 = 2 \quad \beta_0 = -\frac{5}{3}$$

$$f_{LR}(x) = -\frac{5}{3} + 2x^{(1)} + \frac{7}{3}x^{(2)}$$

$$TE_{LR} = \frac{1}{4}\sum_{i=1}^{4}\left(y_i - f_{LR}(x_i)\right)^2$$

$$= \frac{1}{4}\left(\left(2-\frac{8}{3}\right)^2 + \left(5-\frac{14}{3}\right)^2 + \left(1-\frac{2}{3}\right)^2 + \left(5-5\right)^2\right)$$

$$= \frac{1}{4}\left(\frac{4}{9} + \frac{1}{9} + \frac{1}{9}\right) = \boxed{\frac{3}{18}}$$

Exercise 3

a) $k=1$ means that there is only one point to refer to (the closest) for regression. When evaluating the predictor for some $x \in$ Training data, the point to refer to is x itself, because itself is the closest point out of every other point from the data-set. As a result the prediction is totally correct for training points and the error is 0.

** Note that this only applies for a data-set $\{(x_i, y_i)\}_{i=1}^{N}$ where $x_i \neq x_j \ \forall (i,j)$ s.th $i \neq j$.

b|

If the output dimension is 1 and there are 2 points only x_1 and x_2, $\exists f(x) = mx + q$ such that $y_1 = mx_1 + q$ and $y_2 = mx_2 + q$ meaning that the predictor matches the labels of the training points. In plain words, the predictor is the line that passes through x_1 and x_2 and hence, there is 0 distance between x_1 or x_2 and the line. The error is 0.