ASSIGNMENT 4 SOLUTION

– Nayan Man Singh Pradhan

Excercise 1. (Calculating EPE and regressor)

In this task, we revise Examples 3.2 and 3.3 from the lecture notes for a different setup:
Let $X : \Omega \to \mathbb{R}$ be an input variable and $Y : \Omega \to \mathbb{R}$ be an output variable. For the input variable we assume that it follows the uniform distribution as $X \sim \mathcal{U}[-1, 1]$. (Note the other range!) Moreover, we make the very strong assumption to know the "true" dependency between X and Y. Specifically we define Y via

$$Y := g(X), \quad \text{with} \quad g(x) = x^4 .$$

Now we look for a function f that shall approximate that (usually unknown) relationship between X and Y. We claim that

$$f(x) = x^3 ,$$

is a good approximation to $Y = g(X)$.

a) Calculate the expected (squared) prediction error for f. Why is the error that large compared to the lecture example?

b) Find the regressor for $Y = g(x)$.

(4 Points)

a)

$$EPE(f) = E(L_2(Y, f(x)))$$
$$= \int_{\mathbb{R}} \int_{\mathbb{R}} (y - f(x))^2 \cdot p(x, y)\, dx\, dy$$

We know (from knowledge 3.1) that when X, Y is a continuous random variable and Y is given as $Y = g(X)$, where g is a function on the range of X that needs to fulfill some properties. We can give the joint probability distribution density for X and Y by

$$p(x, y) = p_X(x) \cdot \delta(y - g(x))$$

This gives us,

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} (y - f(x))^2 \cdot p_X(x)\, \delta(y - g(x))\, dx\, dy$$
$$= \int_{\mathbb{R}} p_X(x) \int_{\mathbb{R}} (y - x^3)^2\, \delta(y - x^4)\, dy \cdot dx$$

$$\int_{\mathbb{R}} f(y) \cdot \delta(y - y_0)\, dy = f(y_0)$$

$$= \int_{\mathbb{R}} p_X(x) \left(x^4 - x^3 \right)^2 dx$$

$$= \int_{-1}^{1} 1. (x^4 - x^3)^2 \, dx$$

$$= \int_{-1}^{1} (x^8 - 2x^7 + x^6) \, dx$$

$$= \int_{-1}^{1} x^8 \, dx - 2 \int_{-1}^{1} x^7 \, dx + \int_{-1}^{1} x^6 \, dx$$

$$= \left[\frac{x^9}{9} \right]_{-1}^{1} - 2 \times \left[\frac{x^8}{8} \right]_{-1}^{1} + \left[\frac{x^7}{7} \right]_{-1}^{1}$$

$$= \frac{1}{9} + \frac{1}{9} - 2 \times \frac{1}{8} + \frac{1}{8} \times 2 + \frac{1}{7} + \frac{1}{7}$$

$$= \frac{32}{63} \approx 0.5079$$

The error is larger as compared to the lecture because in this question we are working with $g(x)$ with degree 4 and $f(x)$ with degree 3 as compared to the lecture where we were working with $g(x)$ with degree 2 and $f(x)$ with degree 1.

b) We know,

$$p(y|x) = \frac{p(x,y)}{p_X(x)} = \frac{p_X(x) \, \delta(y - g(x))}{p_X(x)} = \delta(y - x^4)$$

Given this, we evaluate the regressor to figure out what would be the best predictor under the given circumstances:

$$E(Y|X=x) = \int_{\mathbb{R}} y \cdot p(y|x) \, dy = \int_{\mathbb{R}} y \cdot \delta(y - x^4) \, dy = x^4 /\!/$$

$$\int_{\mathbb{R}} f(y) \, \delta(y - y_0) \, dy = f(y_0)$$

$\therefore$ The regressor for $Y = g(X)$ is x^4.

Excercise 2. (Regressor derivation)

In the lecture slides, you have seen that the expected (squared) prediction error $EPE(f)$ is given by

$$EPE(f) = E(L_2(Y, f(X))) \,.$$

Theorem 3.1 then states that the function f that minimizes the expected (squared) prediction error $EPE(f)$ is given by

$$f(x) = E(Y|X = x)$$

Note that in the middle of the proof, we encounter an equation

$$EPE(f) = E_X E_{Y|X}[(f(X) - E[Y|X])^2 + 2(f(X) - E[Y|X])(E[Y|X] - Y) + (E[Y|X] - Y)^2|X] \,,$$

where the second term can be dropped. Now, prove that this is possible, i.e. that it holds

$$2E_X E_{Y|X}[(f(X) - E[Y|X])(E[Y|X] - Y)|X] = 0 \,.$$

(4 Points)

Prove that:

$$2 E_X E_{Y|X} [(f(X) - E[Y|X])(E[Y|X] - Y)|X] = 0$$

$$LHS = E_X[2 \cdot (f(x) - E(Y|X))(E[Y|X] - Y)]$$

$$= 2 \int_{\mathbb{R}} (f(x) - E(Y|X))(E[Y|X] - Y) \cdot p_X(x)\, dx$$

$$= 0$$

$$= RHS$$

$\therefore$ LHS = RHS

Proved

$$2 E_X E_{Y|X} [(f(X) - E[Y|X])(E[Y|X] - Y)|X] = 0$$

$$LHS = 2 E_X E_{Y|X} [(f(x) - E[Y|X])(E[Y|X] - Y)|X]$$

$$= E[E[Y|X] - Y)|X]$$

$$= E[Y|X] - E[Y|X]$$

$$= 0$$

$$= 0$$

$$= RHS$$

Proved

Excercise 3. (kNN regression)

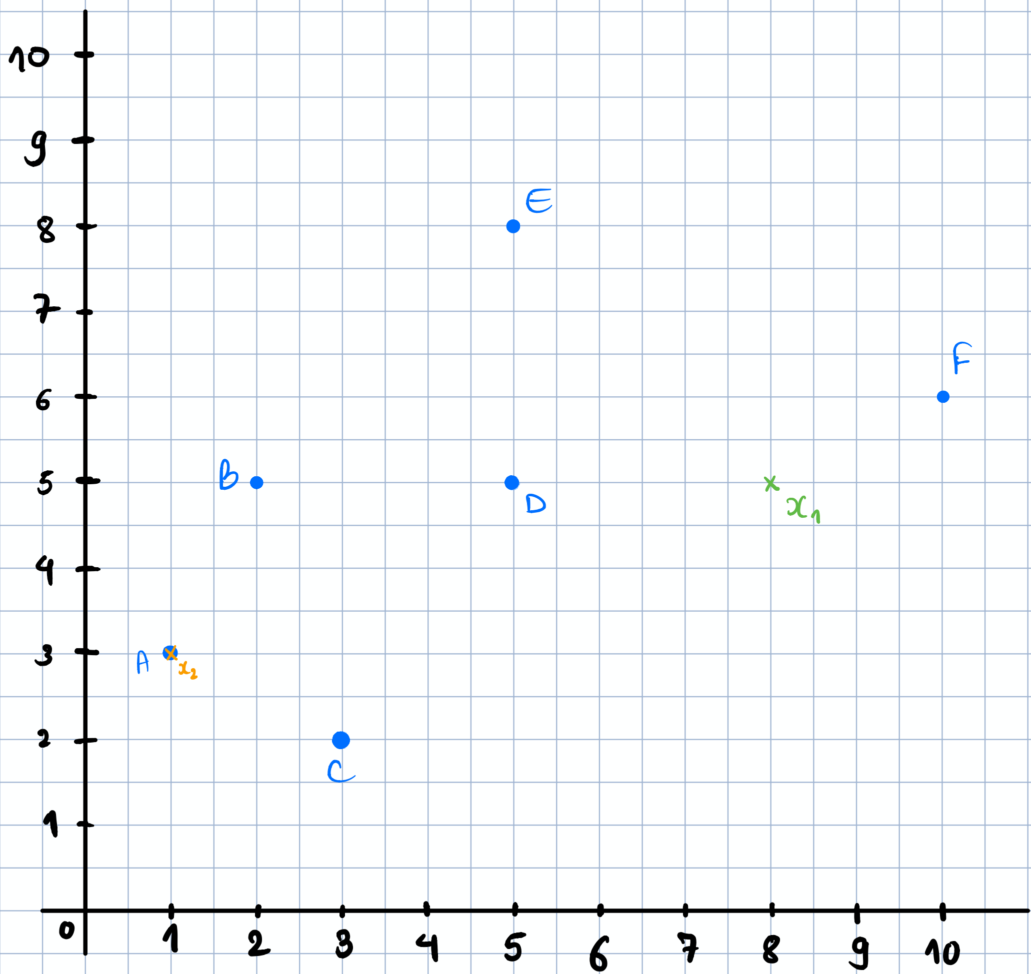
You are given the following training data:

$$\mathcal{T} = \{ (\overset{A}{(1,3)^\top}, 77) , (\overset{B}{(2,5)^\top}, 47) , (\overset{C}{(3,2)^\top}, 55) , (\overset{D}{(5,5)^\top}, 59) , (\overset{E}{(5,8)^\top}, 72) , (\overset{F}{(10,6)^\top}, 60) \}$$

Manually carry a kNN regression prediction for $\mathbf{x}_1 = (8,5)^\top$, $\mathbf{x}_2 = (1,3)$ and $k = 1$, $k = 3$. As part of the task, you have to draw the points in a scatter plot (on paper) and mark the respective neighborhoods that contribute to the final result.

(4 Points)

let, $A(1,3)^T$, $B\ (2,5)^T$, $C\ (3,2)^T$, $D\ (5,5)^T$, $E(5,8)^T$ and $F(10,6)^T$



First, lets calculate the euclidean distance between x_1 and x_2 and all training data points.

$$\mathrm{dist}(A, x_1) = \sqrt{(3-5)^2 + (1-8)^2} = 7.28$$
$$\vdots \qquad \vdots \qquad \vdots$$

I have calculated the distances and illustrated it in Table 1

	A	B	C	D	E	F
x_1	7.28	6.0	5.83	3.0	4.24	2.23
x_2	0.0	2.23	2.23	4.47	6.40	9.48

Table(1).

We know general formula for KNN regression,

$$Y(x_i) = \frac{1}{k}\sum_{j=1}^{k}(y_i)_j \quad \text{where input: } \{(x_i, y_i)\}_{i=1}^{N} \text{ training data}$$

: k neighbor size

When $k=1$,

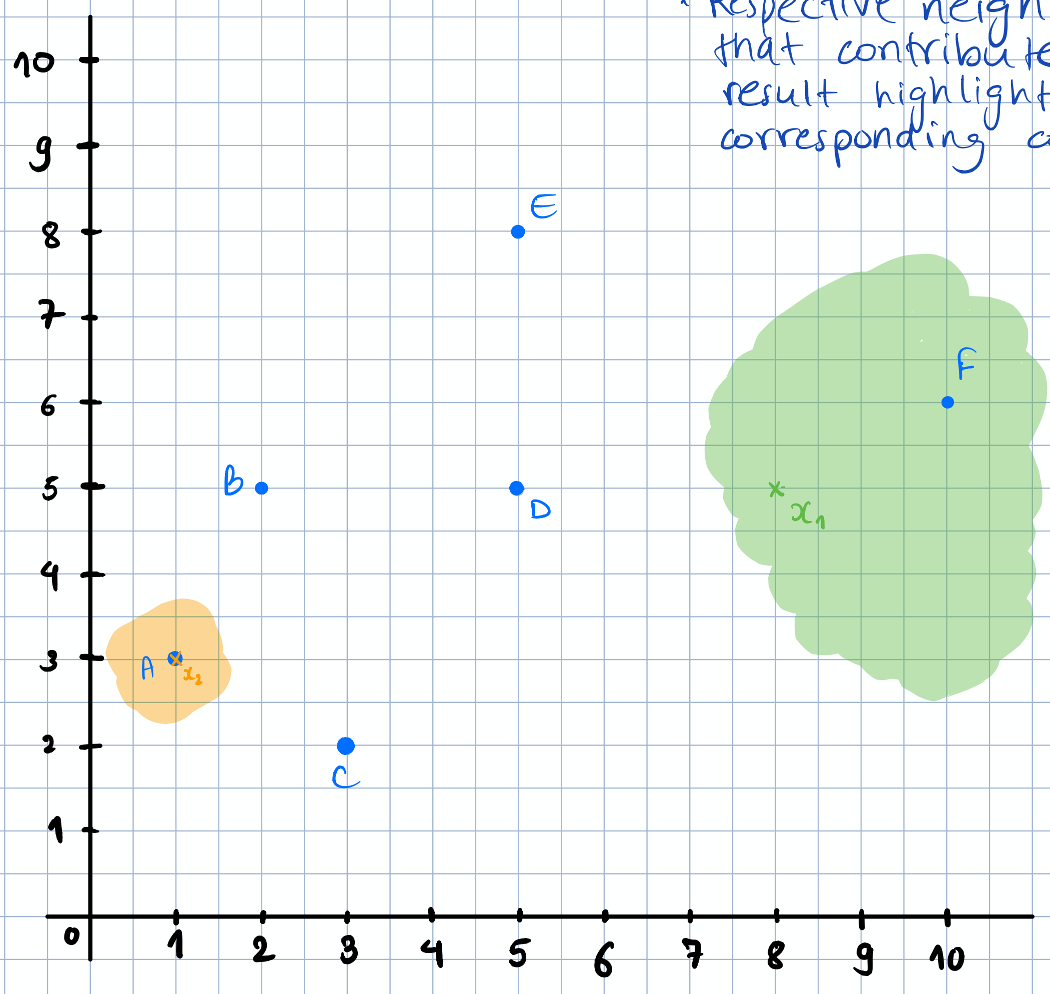
For x_1, the closest neighbor is point F (from Table 1)

$$\therefore Y(x_1) = \frac{1}{1}\sum_{j=1}^{1} 60 = 60 \,//$$

for x_2, the closest neighbor is point A (from Table 1)

$$\therefore Y(x_2) = \frac{1}{1}\sum_{j=1}^{1} 77 = 77 \,//$$

* Respective neighborhood that contribute to final result highlighted with corresponding color



When $K = 3$,

For x_1, the closest neighbors are F, D and E (from Table 1)

$$\therefore Y(x_1) = \frac{1}{3}(60 + 59 + 72) = 63.67$$

for x_2, the closest neighbors are A, B and C (from Table 1)

$$\therefore Y(x_2) = \frac{1}{3}(77 + 47 + 55) = 59.67$$

* Respective neighborhood that contribute to final result highlighted with corresponding color