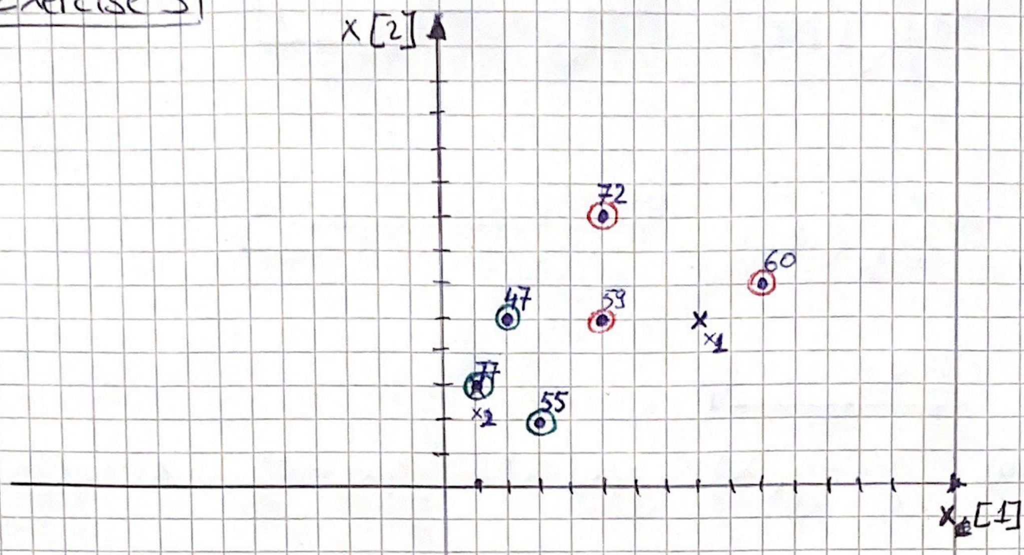
Exercise 3



- Case 1 : $k=1$. Let f_{k_1} be the KNN regressor with $k=1$.

→ The first point of the training set coincides with the test point x_2. Thus, since $k=1$, there's no need to look further. $f_{k_1}(x_2) = 77$

→ By looking at the plot by naked eye, the fourth and sixth points of the training set seem nearest to x_1. $\left\| x_1 - \begin{pmatrix}5\\5\end{pmatrix} \right\|$ compared to $\left\| x_1 - \begin{pmatrix}10\\6\end{pmatrix} \right\|$.

$\sqrt{(8-5)^2+(5-5)^2} = 3$ compared to $\sqrt{(8-10)^2+(5-6)^2}$.

Clearly, $\sqrt{5} < 3$. $\begin{pmatrix}10\\6\end{pmatrix}$ is the nearest point to x_1. Hence, $f_{k_1}(x_1) = 60$.

- Case 2 : $k=3$. Let f_{k_2} be the KNN regressor with $k=3$.

The three nearest points to x_1 have been marked in red. The three nearest points to x_2 have been marked in green.

$\rightarrow f_{k_2}(x_2) = \dfrac{77+47+55}{3} = \boxed{59.7}$

$\rightarrow f_{k_2}(x_1) = \dfrac{60+72+59}{3} = \boxed{63.7}$

Exercise 1

a) $Y := g(X) \quad g(x) = x^4 \quad f(x) = x^3 \quad X \sim \mathcal{U}[-1,1]$

$$EPE(f) = \mathbb{E}[L_2(Y, f(X))]$$

$$EPE(f) = \iint_{R_x R_y} (y - f(x))^2 \rho(x,y)\, dx\, dy =$$

$$\iint_{R_x R_y} \underbrace{\overbrace{(y - f(x))^2}^{A}}_{g(y)} \rho_X(x) \overbrace{\delta(y - \underbrace{x^4}_{y_0})}^{B} dx\, dy =$$

STEP EXPLANATION

$$\int_{R_y} A \cdot B\, dy = \int_{R_y} g(y)\,\delta(y - y_0)\, dy = g(y_0) = (x^4 - f(x))^2$$

*Please note that $\rho_X(t) = \frac{1}{2}\ \forall t \in [-1,1]$ and $\rho_X(t) = 0$ otherwise because $X \in \mathcal{U}[-1,1]$

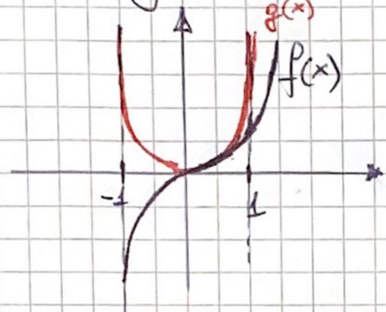
$$\int_{R_x} (x^4 - x^3)^2 \rho_X(x)\, dx =$$

$$\frac{1}{2}\int_{R_x} (x^4 - x^3)^2\, dx = \frac{1}{2}\int_{-1}^{1} (x^4 - x^3)^2\, dx = \frac{1}{2}\int_{-1}^{1} x^8 - 2x^7 + x^6\, dx =$$

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

$$EPE(f) = \boxed{0.254}$$

The error is notably higher compared to the lecture's example because $\forall x \in [-1, 0]$ $f(x)$ diverges to the third quadrant of the Cartesian plane while $g(x)$ is decreasing monotonously in the second quadrant.



b) The regressor for $Y = x^2$

$$\mathbb{E}[Y|X=x] = \int_{R_y} y\, p(y|x)\, dy = \int_{R_y} y \frac{p(y,x)}{p_X(x)}\, d\cancel{y}$$

From the definition of $p(y,x)$ using the Dirac function, we have:

$$\mathbb{E}[Y|X=x] = \int_{R_y} y \cdot \frac{\cancel{p_X(x)}\,\delta(y-x^4)}{\cancel{p_X(x)}}\, d\cancel{y}$$

$$\Rightarrow \mathbb{E}[Y|X=x] = \underbrace{\int_{R_y} \underset{f(y)}{Y}\, \delta(y - \underset{y_0}{x^4})\, dy}_{\text{From Knowledge 3.1 of the Lecture notes}} = \boxed{x^4}$$