Question 1

In this question you will be considering three different regression techniques \((x \in \mathbb{R}^N \text{ and } y \in \mathbb{R})\), namely:

1. **Regular Least Squares**
   
   - regressor: \( y = W^T x + b \)
   - optimisation: \( W = (X^T X)^{-1} X^T y \)

2. **Weighted Least Squares**
   
   - regressor: \( y = W^T x + b \)
   - optimisation: \( W = (Z^T Z)^{-1} Z^T y \) where \( Z = BX^T \)

3. **Locally Weighted Regression**
   
   - regressor: \( y = \left( \sum_{i=1}^M \beta_i(x) \cdot y^i \right) / \left( \sum_{i=1}^M \beta_i(x) \right) \)
   
   The beta is a kernel density function centred on a point \( i: \beta_i(x) = \exp(-\frac{1}{2}||x^i - x||) \)
   
   - optimisation: no-optimisation, data driven.

In figure 1, three different data sets are given

![Data sets](image)

Figure 1: Three data sets, the black circles depict the data points. \( x \) is the input and \( y \) is the output and we wish to estimate \( y = f(x) \).

1. Draw the solution that RLS would give you for data sets 1 to 3 (do not consider the colored points in dataset 2).
2. Given this set of weights, $\beta = \left[ \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 0, \frac{1}{4} \right]$, apply Weighted Least Squares to data set 1 and draw the resulting regression function.

3. What solution WLS would give for data set 2, considering first the blue (point $x^1$) and then the red data point (point $x^2$). Use $\beta(x) = \frac{1}{x}$

4. Draw the solutions of LWR for data set 3 with each of the given kernels in Fig 1, Bottom right.

**Question 2**

a) Fig. 2 below shows a 1-dimensional SVR problem with 2 data points. For which parameters would the $\epsilon$-SVR method fail to find a non-trivial solution? Draw the trivial solution that would result from one such choice of parameters. Also, show by drawing, the effect of changing kernel width for both the trivial and non-trivial solutions.

![Figure 2: $\epsilon$-SVR, case a](image)

b) Fig. 3 shows a 1-dimensional SVR problem with 2 data points that are superimposed on x (i.e. they have the same value for x but a different value for y). Discuss the effect of the $\epsilon$ parameter on the regression function in this case.

![Figure 3: $\epsilon$-SVR, case b](image)
Question 3 (Do at home)

In the lecture you have covered linear function estimators of the following form:

$$y^i = f(x^i; w, b) = w^T x^i + b$$  \hspace{1cm} (1)

Where $w \in \mathbb{R}^N$ and $x \in \mathbb{R}^N$ are $(N \times 1)$ column vectors, $b$ is the scalar intercept and $y$ is the predictor. Given you have a set of $M$ data points, $X = [x^1, \ldots, x^i, \ldots, x^M]$, and associated predictors, $y = [y^1, \ldots, y^i, \ldots, y^M]$. Consider the Sum of Squares Error (SSE)$^1$ as your loss function and derive the optimal choice of parameters of the linear regressor for the **Bivariate** case:

$$y^i = w \cdot x^i + b$$  \hspace{1cm} (2)

$$SSE = \sum_{i=1}^{M} (y^i - f(x^i))^2 = \sum_{i=1}^{M} e_i^2$$  \hspace{1cm} (3)

Where $e_i$ is the error between the target and predicted value.

---

$^1$Sum of Squares Error (SSE) is a loss function which quantifies the penalty/cost off badly predicting the target values, $y$. This is very similar to classification, were $y^i \in \mathbb{Z}$ where discrete, but now for regression they are continuous, $y^i \in \mathbb{R}^1$. 