The best $k$ is 23 with accuracy 81.23%. The result is based on Euclidean distance. Any solution with a similar graph was accepted as a correct solution.
The best $\lambda$ is 1.2 with value 2.1243. $\lambda \in [0.9, 1.5]$ were all accepted as correct solutions.
Question 3

Part (a)

Write the objective in matrix form

\[ L(w, b) = \sum_{n=1}^{N} r_n (y_n - w^T x_n + b)^2 \]

\[ = (y - v^T \bar{X}) R (y - v^T \bar{X})^T \]

\[ = y R y^T - 2y R \bar{X}^T v + v^T \bar{X} R \bar{X}^T v \]

where

\[ v = \begin{bmatrix} -b \\ w \end{bmatrix} \]

\[ y = [y_1, \ldots, y_N] \]

\[ \bar{X} = \begin{bmatrix} 1 \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots 1 \\ x_1, \ldots, x_N \end{bmatrix} \]

\( R \) is a diagonal matrix with entries \( \text{diag}(r_1, r_2, \ldots, r_N) \)

Assuming the weights \( r_n \) are positive, the objective is a convex quadratic function of \( w \) and \( b \). So, the minimum occurs at the point of zero gradient:

\[ 0 = \nabla L(v) = -2 \bar{X} R y^T + 2 \bar{X} R \bar{X}^T v \]

Solve for \( v \) and we get

\[ v = (\bar{X} R \bar{X}^T)^{-1} \bar{X} R y^T \]

Part (b)

Suppose \( y_n = w^T x_n - b + \epsilon_n \) where \( \epsilon_n \) is a zero-mean Gaussian random variable with variance \( \sigma_n^2 \). Thus, the conditional density of \( y_n \) given \( x_n \) is Gaussian with mean \( w^T x_n - b \) and variance \( \sigma_n^2 \). The negative log-likelihood is

\[ -l(w, b) = -\log P\{y_1, \ldots, y_N|x_1, \ldots, x_N\} \]

\[ = -\log \prod_{n=1}^{N} P\{y_n|x_n\} \]

\[ = -\sum_{n=1}^{N} \log P\{y_n|x_n\} \]

\[ = \sum_{n=1}^{N} \log(\sqrt{2\pi\sigma_n^2}) + \sum_{n=1}^{N} \frac{(y_n - w^T x_n + b)^2}{2\sigma_n^2} \]

The first term is constant with respect to \( w \) and \( b \). So minimizing the negative log-likelihood is equivalent to minimizing only the second term, which is the same as our objective function with \( r_n = \frac{1}{\sigma_n^2} \). The variance of measurement \( n \) in this model is thus inversely proportional to the weight \( r_n : \sigma_n^2 \propto \frac{1}{r_n} \).