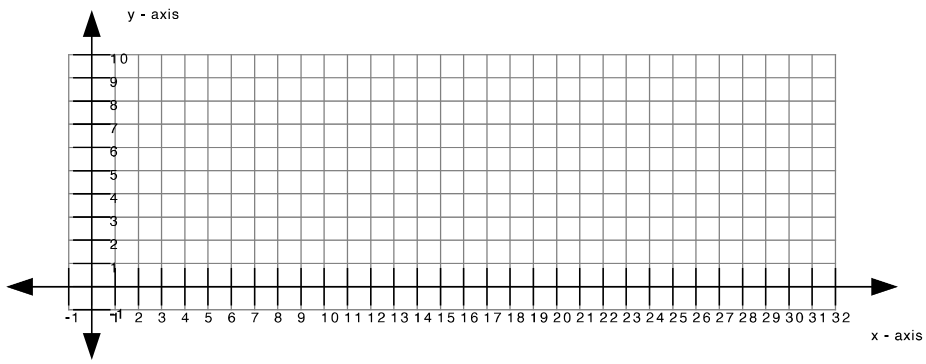
4. a) K-NN Classification and Regression

i) Compare K-NN classifier with SVM classifier (in the original data domain) in terms of training complexity, classification boundaries, and the complexity of prediction of a new instance. [30 marks]

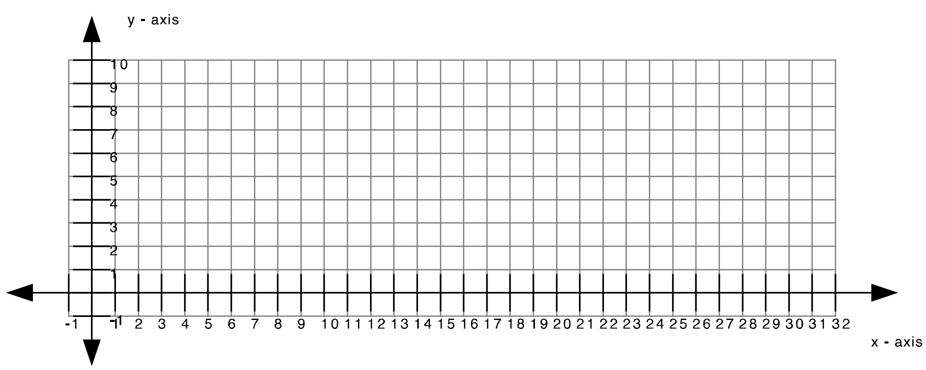
ii) You are given the following dataset of (x, y) pairs:

$$(2,2),(6,4),(10,6),(16,6),(18,0),(24,4)$$

Assume that we apply K-NN regression (using Euclidean distance) on this dataset, and let $f(x)$ denote the resulting K-NN regression function that maps new instances to predicted values. Plot function $f(x)$ for $K = 1$ for the range of $x \in [-1, 32]$. [10 marks]



iii) Plot the K-NN regression function $f(x)$ for the same dataset for $K = 2$ (with uniform averaging) over the same range. [20 marks]



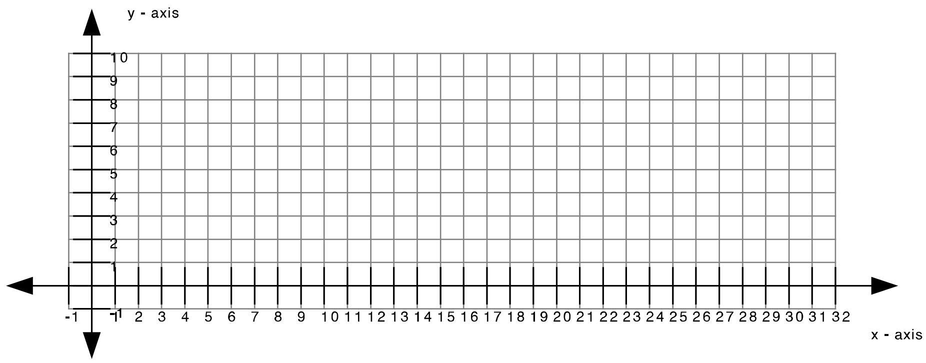
iv) Apply weighted averaging for $K = 2$ with weights proportional to the distance between the new instance and the two nearest points. Let

$\mathrm{nn}_k(x)$ denote the index of the k-th nearest point in the dataset to point x (as defined in the lecture), and $d_k(x)$ denote the Euclidean distance between x and $x_{\mathrm{nn(x)}}$. Then, we set the prediction corresponding to x as

$$\hat{y} = f(x) = \frac{d_2(x)}{d_1(x)+d_2(x)} y_{\mathrm{nn}_1(x)} + \frac{d_1(x)}{d_1(x)+d_2(x)} y_{\mathrm{nn}_2(x)}.$$

Plot the regression function $f(x)$ with the above weighted averaging for the same dataset.

[30 marks]



v) Compare the mean-squared training error between the three regression functions above. [10 marks]