Assignment 1

Question 1

Assumption: The hypothesis class is an axis-aligned rectangle.

In the case of a single rectangle, the positive data points shall form one cluster, while with two or more rectangles (m>1), we can have separate or overlapping clusters in the hypothesis space.

In the worst-case scenario,

$$m \text{ (number of clusters or rectangles)} = n \text{ (number of positive instances)}$$

For 2 rectangles the VC dimension will be 8 and VC dimension for m=n rectangles will be 4m=4n.

Therefore, the VC dimension for hypothesis with rectangles m>1 lies within,

$$\boxed{8 \leq VC\ dimension \leq 4m = 4n}$$

Question 2

In MSE (mean squared error) & RMSE (root mean squared error), it squares the differences deviating the fitted line producing significant errors. We will also find the error distribution curve skewed. MSE & RMSE penalizes large errors. Instead, we can use MAE (Mean Absolute Error), where we take the mean of the absolute difference, resulting in better resilience to large errors than MSE and RMSE.

$$MAE \text{ (Mean Absolute Error)} = \frac{1}{N}\sum_{t=1}^{N}|r^t - g^t(x)|$$

Question 3

Mean squared error (MSE) for first-order function

$$g(x) = w_1 x + w_0$$

$$MSE = E(w_1, w_0 \mid X) = \frac{1}{N}\sum_{t=1}^{N}[r^t - g^t(x)]^2 = \frac{1}{N}\sum_{t=1}^{N}[r^t - (w_1 x^t + w_0)]^2$$

Minimizing the loss function by taking the partial derivatives of E wrt. parameters and equating them to zero.

$$\frac{\partial E}{\partial w_0} = \frac{1}{N}\sum_{t=1}^{N} 2[r^t - (w_1 x^t + w_0)] = \sum_{t=1}^{N}[r^t - (w_1 x^t + w_0)] = 0$$

$$\sum_t r^t = w_1 \sum_t x^t + N w_0$$

$$\boxed{w_0 = \frac{1}{N}\sum_t r^t - \frac{w_1}{N}\sum_t x^t = \bar{r} - w_1 \bar{x}}$$

$$\frac{\partial E}{\partial w_1} = \frac{1}{N}\sum_{t=1}^{N} 2[r^t - (w_1 x^t + w_0)]x^t = \sum_{t=1}^{N}[r^t - (w_1 x^t + w_0)]x^t = 0$$

$$\sum_t r^t x^t = \sum_t w_1 {x^t}^2 + w_0 \sum_t x^t$$

$$\sum_t r^t x^t = \sum_t w_1 {x^t}^2 + (\bar{r} - w_1 \bar{x}) \sum_t x^t$$

$$\sum_t r^t x^t = w_1 \left(\sum_t {x^t}^2 - \bar{x}\sum_t x^t\right) + \bar{r}\sum_t x^t$$

$$\sum_t r^t x^t = w_1 \left(\sum_t {x^t}^2 - \bar{x}N\bar{x}\right) + \bar{r}N\bar{x}$$

Where $\bar{r} = \frac{1}{N}\sum_t r^t$ and $\bar{x} = \frac{1}{N}\sum_t x^t$

$$\boxed{w_1 = \frac{\sum_t r^t x^t - N\bar{r}\bar{x}}{\sum_t {x^t}^2 - N\bar{x}^2}}$$

Question 4

For a second-order equation, to estimate the parameters, we find the minima of the MSE loss function by taking a partial derivate of the error function wrt. w_2, w_1 & w_0 and setting it to zero. The same steps we followed in Question 3. Including an optimum number of parameters coupled with the best minimizing error function can reduce the error.

$$g(x) = w_2x^2 + w_1x + w_0$$

We will root mean squared error as loss function for the estimation,

$$E(w_2, w_1, w_0 \mid X) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[r^t - g^t(x)]^2}$$

We minimize the error by taking the partial derivatives error function for all the variables and setting it to zero,

$$\frac{\partial E}{\partial w_2} = 0 \ \& \ \frac{\partial E}{\partial w_1} = 0 \ \& \ \frac{\partial E}{\partial w_0} = 0$$

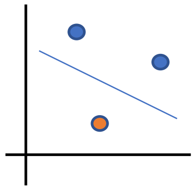
Solving for the above gives estimated values for w_2, w_1, and w_0.

Question 5

To prove: For a hypothesis class, set of lines. VC dimension for a line is three.

Labeling for N=3 points are 8. Blue dots are positive classes, and red is negative classes.

For all combinations of three points (excluding consecutively placed points), there exists a line that can separate both classes. Now VC dimension is greater than or equal to 3.



Let us consider a hypothesis space with four points. A combination exists where two classes cannot be shattered with a line. Therefore, the VC dimension of a line is three.