Theodore Kim
CS-UY 4563-A: Introduction to Machine Learning
Homework #3

1. Determine if the pairs of true functions ($f(x)$) and model functions ($g(x)$) are (i) linear and (ii) underfitted, and (iii) if not underfitted, what is the true parameter?

a. $f(x) = 1 + 2x$ **and** $h(x) = w_0 + w_1x + w_2x^2$

(i) The model class **is** linear

(ii) The model function is **not underfitted** to the true function

(iii) the true parameters are: $w_0 = 1, w_1 = 2, w_2 = 0$

b. $f(x) = 1 + x + 3x^2 + 4x^3$ **and** $h(x) = w_0 + w_1x + w_2x^2$

(i) the model **is** linear

(ii) the model function **is underfitted** to the true function

c. $f(x) = (x_1 - x_2)^2 = (x_1^2 - 2x_1x_2 + x_2^2)$ **and** $h(x) = w_0 + w_1x_1 + w_2x_2 + w_3x_1^2 + w_4x_2^2$

(i) the model **is not** linear

(ii) the model function **is not underfitted**

(iii) the true parameters are: $w_0 = 0, w_1 = x_2, w_2 = x_1, w_3 = 1, w_4 = 1,$

2. Cancer diagnostic test for 3 models: (1) diagnostic measurement depends linearly only on cancer volume.

a. Cancer volume (x) and age (x_1) are both continuous numerical amounts and cancer type (x_2) would be assigned 0 or 1 for Types 1 or 2 respectively.

Model 1: $\hat{y} = w_0 + w_1x_1$

Model 2: $\hat{y} = w_0 + w_1x_1 + w_2x_2$

Model 3: $\hat{y} = w_0 + w_1x_1 - w_2(x_2 - 1) + w_3x_2$ *(There are two different weights for each type)*

b. Model 1 has one parameter (cancer volume) and model 2 has two parameters (cancer volume and age). Model 2 is more complex as it has more parameters

c. **Model 1:** $X = \begin{bmatrix} 0.7 \\ 1.3 \\ 1.6 \end{bmatrix}$

Model 2: $X = \begin{bmatrix} 0.7 & 55 \\ 1.3 & 65 \\ 1.6 & 70 \end{bmatrix}$

Model 3: $X = \begin{bmatrix} 0.7 & 55 & 0 \\ 1.3 & 65 & 1 \\ 1.6 & 70 & 1 \end{bmatrix}$

d. Based upon the results of the ten-fold cross validation, model 3 should be utilized as it has the lowest MSE for both the training and test sets

3. Suppose below training models:

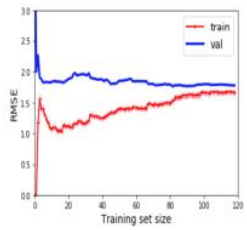

Figure 1: A

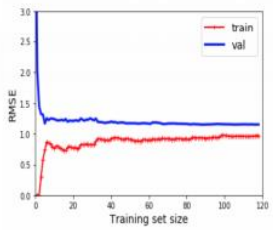

Figure 2: B

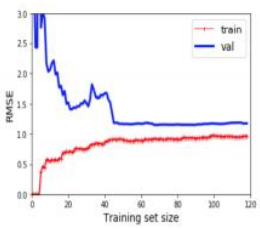

Figure 3: C

Model A: Model A seems to be mostly overfit for small sizes of the training set as there is high variance (difference between the validation and training set) and low bias (low error for the training set) and underfit for larger training set sizes as there is a small variance between the two sets but high bias as there is high error for both.

Model B: Model B seems like it has a lower overall variance as both sets maintain a closer error more constantly with fluctuations in the training set size. Given this and the low mean squared error for both sets, this model is less likely to be either over or underfitted compared to the other two models.

Model C: Model C is mostly overfit, given the high variance between the sets and low bias of the training set model.

All models become more overfit as the training set size increases as there is more information available to the model to train on