ECE421

2021

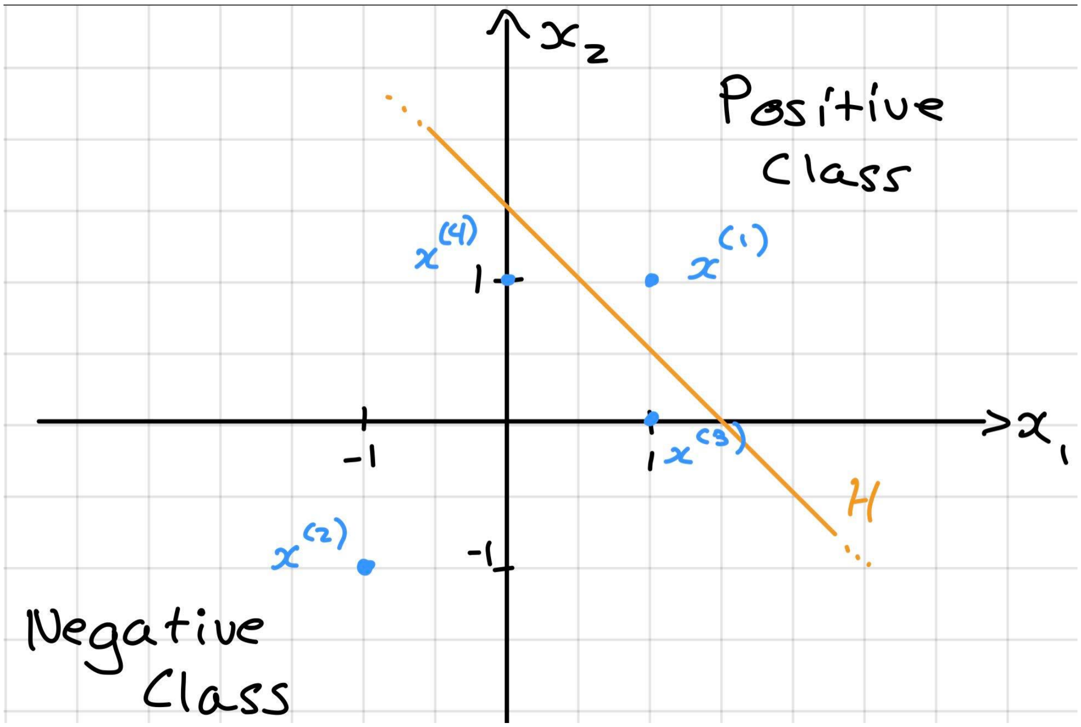
Assignment Three

Chase McDougall

Engineering Science

Problem 1

1.



Equation for the maximum margin hyperplane:

$$x_2 = 1.5 - x_1$$

Or

$$z = \vec{w}^T\vec{x} + b$$

$$Z = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 1.5$$

Where the sign of Z indicates the class which the given sample falls in.

2.

$x^{(1)}$, $x^{(3)}$, and $x^{(4)}$

3.

$$\min_{\vec{w},b} \frac{1}{2}||\vec{w}||^2 \; s.t. \; (\vec{w} \cdot \vec{x}^{(j)} + b)y_j \geq 1, \; \forall j$$

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\vec{w} \cdot \vec{w} - \sum_j \alpha_j[(\vec{w} \cdot \vec{x}^{(j)} + b)y_j - 1]$$

4.

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\vec{w} \cdot \vec{w} - \sum_j \alpha_j[(\vec{w} \cdot \vec{x}^{(j)} + b)y_j - 1]$$

$$\frac{\partial}{\partial \vec{w}}(L(\vec{w}, b, \vec{\alpha})) = \frac{\partial}{\partial \vec{w}}(\frac{1}{2}\vec{w} \cdot \vec{w} - \sum_j \alpha_j[(\vec{w} \cdot \vec{x}^{(j)} + b)y_j - 1])$$

Performing this partial derivative, setting it to zero and rearranging will provide the following expression (See Problem 1-4 Steps for steps):

$$\vec{w} = \sum_j \alpha_j y_j \vec{x}^{(j)}$$

$$\frac{\partial}{\partial b}(L(\vec{w}, b, \vec{\alpha})) = \frac{\partial}{\partial b}(\frac{1}{2}\vec{w} \cdot \vec{w} - \sum_j \alpha_j[(\vec{w} \cdot \vec{x}^{(j)} + b)y_j - 1])$$

Repeating the same steps, but now we take the partial derivative with respect to b gives the following (See Problem 1-4 Steps for steps):

$$0 = \sum_j a_j y_j$$

5.

For this question we will using the following three equations form the previous parts:

$$L(\vec{w}, b, \vec{\alpha}) = \frac{1}{2}\vec{w} \cdot \vec{w} - \sum_j \alpha_j[(\vec{w} \cdot \vec{x}^{(j)} + b)y_j - 1]$$

$$\vec{w} = \sum_j \alpha_j y_j \vec{x}^{(j)}$$

$$0 = \sum_j a_j y_j$$

By substituting the second and third equations into the first and rearranging we can arrive at the following (Derivation available in Problem 1-5 Steps):

$$\max_{\vec{a}} \sum_j \alpha_j - \frac{1}{2}\sum_j \sum_k y_j y_k \alpha_j \alpha_k (\vec{x}^{(j)} \cdot \vec{x}^{(k)})$$

6.

Because $\vec{x}^{(2)}$ is not a support vector it's associated dual variable is zero. Ie:

$$\alpha_2 = 0$$

7.

Making use of the following equation:

$$0 = \sum_j a_j y_j$$

We know the values of y_j to be the following (given in the problem):

$$y_2 = y_3 = y_4 = -1$$
$$y_1 = 1$$

Plugging these into the summation we get:

$$\alpha_1 = \alpha_2 + \alpha_3 + \alpha_4$$
$$\Rightarrow \alpha_1 = \alpha_3 + \alpha_4$$

8.

For this problem we will use the equality found in part 7 in conjunction with the following equation.

$$\max_{\vec{a}} \sum_j \alpha_j - \frac{1}{2} \sum_j \sum_k y_j y_k \alpha_j \alpha_k (\vec{x}^{(j)} \cdot \vec{x}^{(k)})$$

Substituting in the previous equality for all values of α_1, taking the derivative with respect to both α_3 and α_4 and setting these to zero will provide the following values:

$$\alpha_3 = \alpha_4 = 2$$
$$\alpha_1 = \alpha_3 + \alpha_4 = 4$$

Full steps for this process can be found in Problem 1-8 Steps.

9.

Using the following relation from question 4 we are able to expand and derive the value of $\vec{w}$.

$$\vec{w} = \sum_j \alpha_j y_j \vec{x}^{(j)}$$

Substituting in all our known values provides the following (Full Solution available in Problem 1-9 Steps):

$$\vec{w} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

10.

We know that when we are looking at the following inequality, the inequality becomes an equality for support vectors.

$$\min_{\vec{w},b} \frac{1}{2}||\vec{w}||^2 \; s.t. \; (\vec{w} \cdot \vec{x}^{(j)} + b)y_j \geq 1, \; \forall j$$

This allows us to obtain the following:

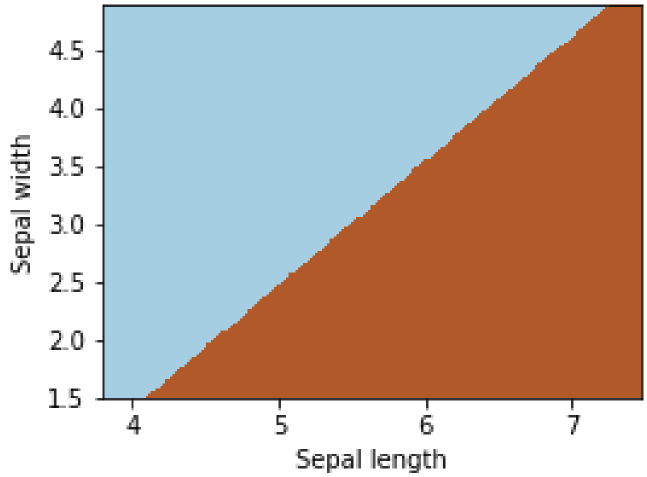
$$(\vec{w} \cdot \vec{x}^{(1)} + b)y_1 = 1$$

$$\begin{bmatrix} 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + b = 1$$

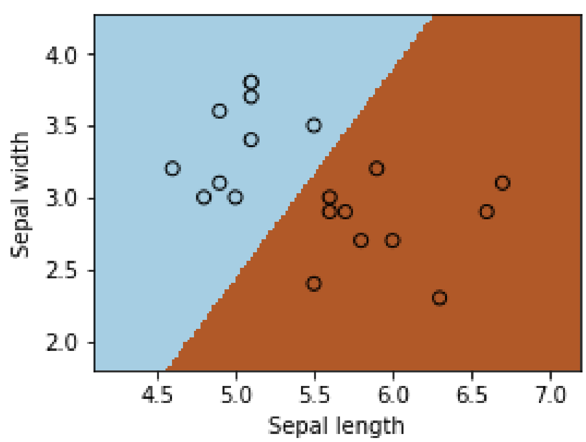
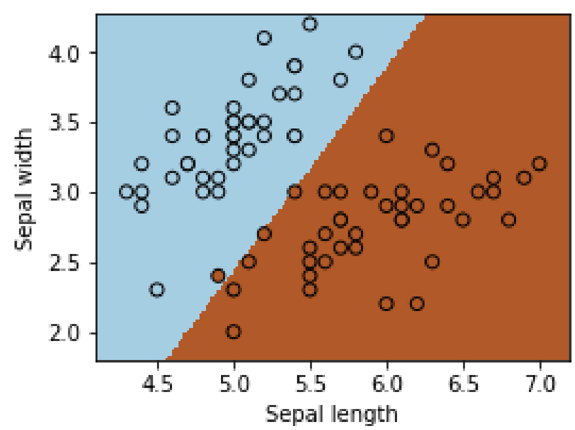
$$\Rightarrow b = -3$$

Problem 2

1.



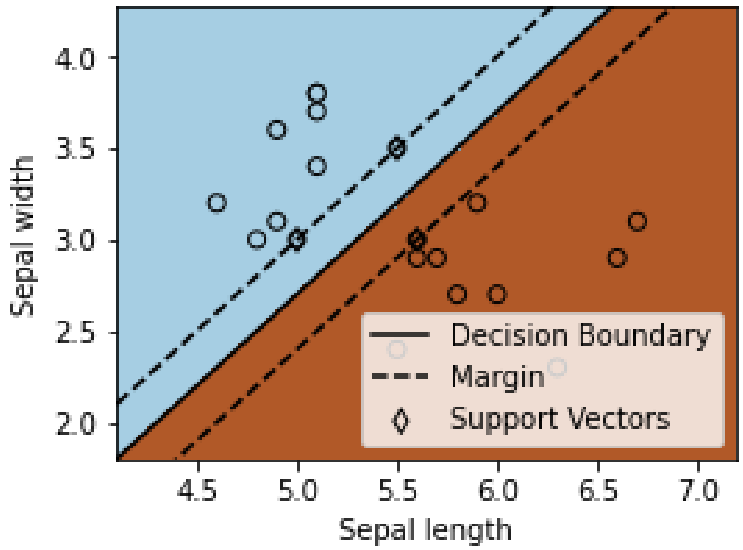
2.

	
Training points	Test Points

My implemented binary linear classifier is accurate in classifying the 20 training points, with all being correctly classified.
However, the test set has a couple points straddling the decision boundary and a point clearly misclassified giving an accuracy <100% and equal to 98.75%.

Therefore, this model is able to accurately predict most data points, but has been shown to mislabel some data points within the current training set.

3/4.



The above plot depicts the hard margin SVM classifier of the Iris sklearn dataset. Plotted in the above will be 20 of the first 100 data points, which makeup our training set. Our **three** Support Vectors are indicated by a diamond shape.

In order to understand how to identify support vectors we can look back to our approach in problem 1. The following equation which was introduced in Problem 1 turns into an equality when we are dealing with support vectors. This condition allows us to check if a given point is a support vector.

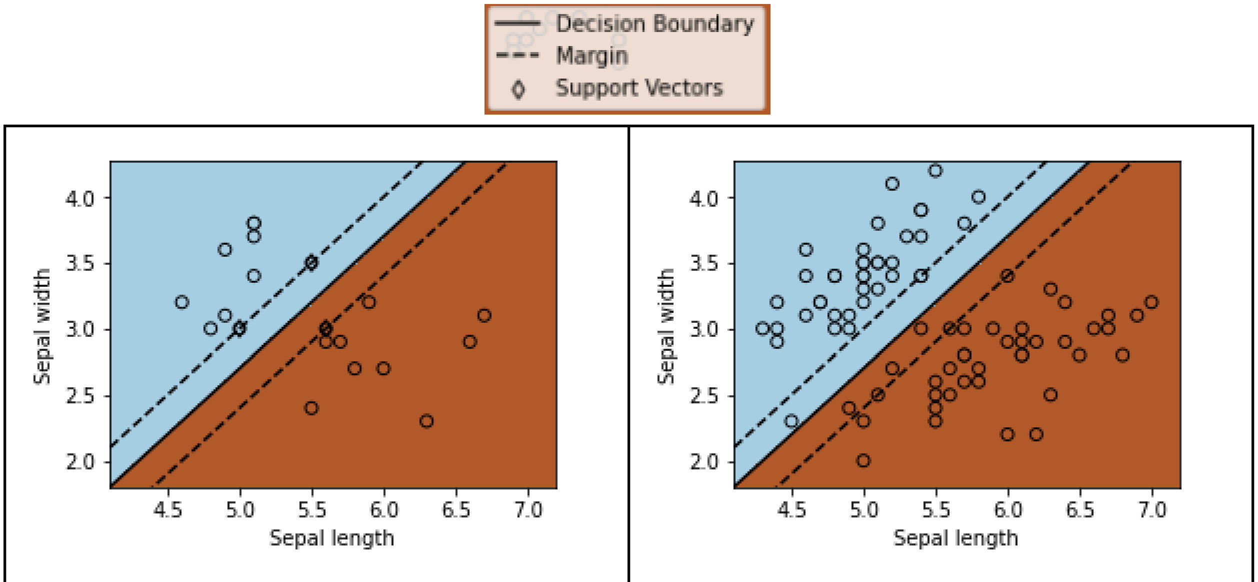
$$\min_{\vec{w},b} \frac{1}{2}||\vec{w}||^2 \; s.t. \; (\vec{w} \cdot \vec{x}^{(j)} + b)y_j \geq 1, \; \forall j$$

For example the following represents the Support Vector located at $\begin{bmatrix} 5 \\ 3 \end{bmatrix}$:

$$\begin{bmatrix} -3.33 & 3.33 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \end{bmatrix} + (2.299)(3.33) \approx 1$$

5.

Plot with test_size=0.8



Training Points	Test Points

From the above two plots we can see that all of our training and test points have been correctly classified. While the training points are more confidently placed within their respective classifications, the test points were still correctly classified. Therefore the SVM method is a more accurate classifier than the Binary Linear Classifier.

6. Margin value is equal to the total orthogonal distance between the margins, denoted M:

$$M = 2(d) = 2\frac{1}{||\vec{w}||_2} = \frac{2}{\sqrt{3.333^2 + (-3.333)^2)}} \approx 0.424$$

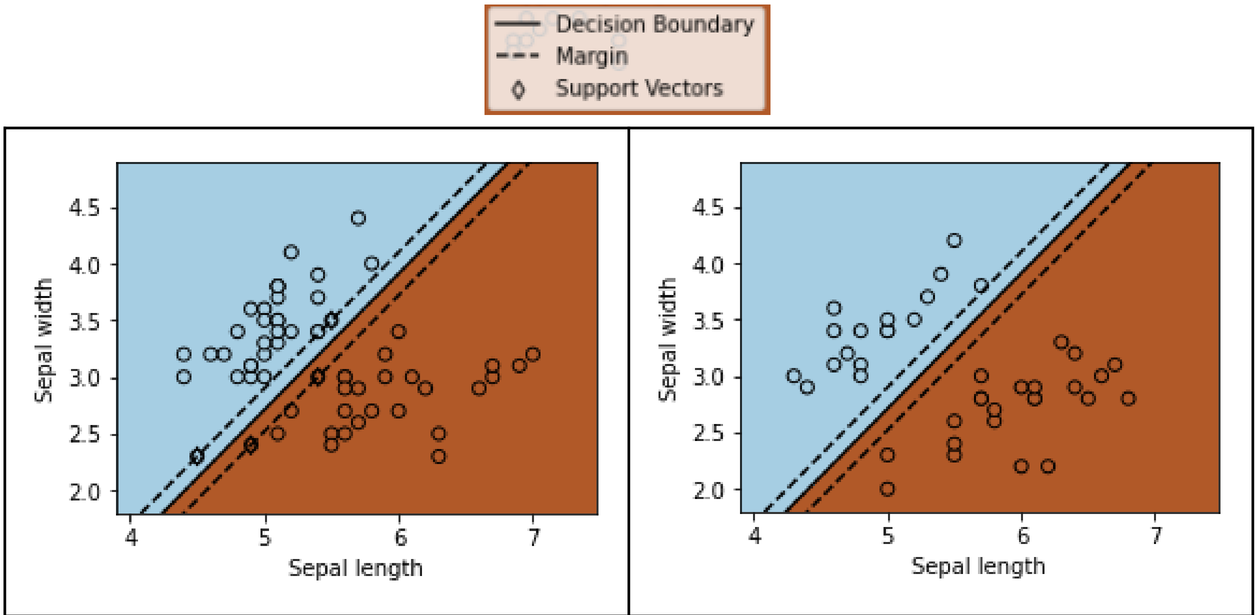
Where d is the orthogonal distance from the decision boundary to a margin.

7.

$\vec{w}$ is orthogonal to the decision boundary.

8.

Plot with test_size=0.4

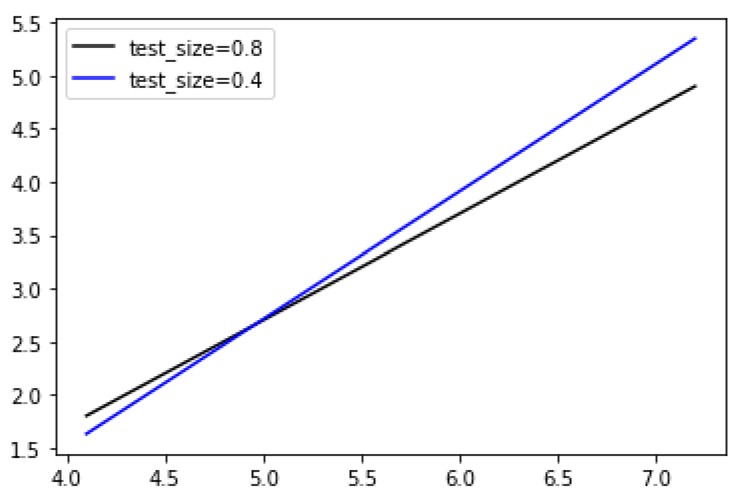


Training Points	Testing Points

While we are still correctly classifying all points, we can see that there are no longer any testing points falling within our margin. If we define the accuracy of our model to be:

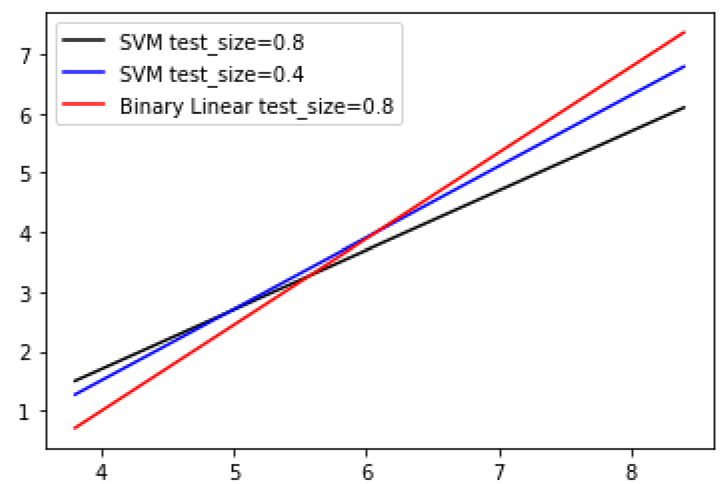
$$\frac{correct\ predictions}{total\ predictions}$$

Then we find that the accuracy of a test_size of 0.8 and 0.4 are equivalent based on our testing data sets, as they both correctly classify 100% of the testing points.



The plot above shows the decision boundaries of both test_size cases. From the plot we can see that there is clearly a change. This is because when changing the test_size from 0.8 to 0.4, we significantly increased the amount of training data available for our model. This increase meant that we found support vectors, which were closer to the decision boundary than our early support vectors. These new closer support vectors became the defining vectors for our new decision boundary.

9.



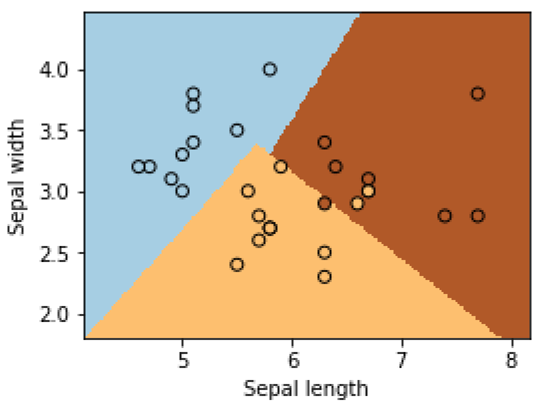
Decision Boundaries of Various Models

No. As per the above plot, we see that the Binary Linear Classifier's decision boundary is different from both SVM test size's decision boundaries.

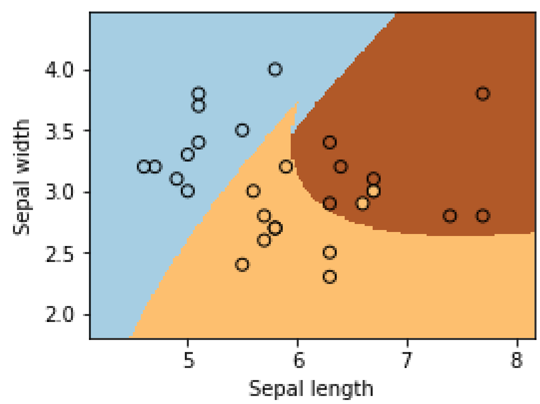
10.

Due to the addition of a third class, we can no longer separate the classes with a single linear division. Therefore there are a couple approaches we can take:

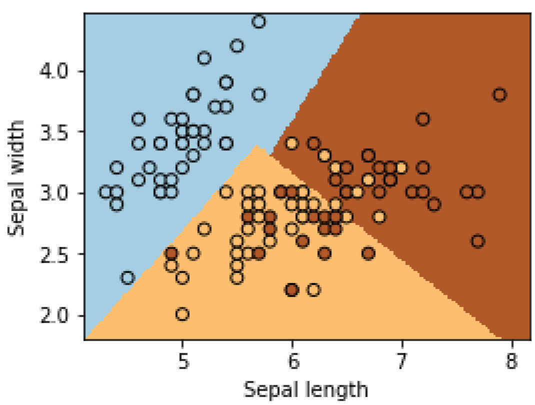
1. Linear:
   We could simply use an additional linear separator to create another division and three "zones".
2. Polynomial:
   For some datasets a polynomial fit may be more suitable, as it could lead to an increased accuracy when predicting classes.



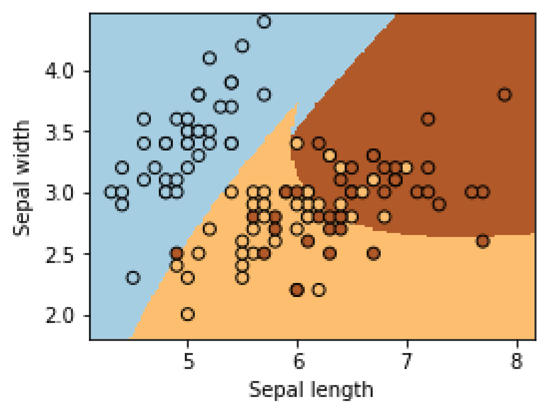
Linear Fit (Training Set)



Polynomial Fit (Training Set)



Linear Fit (Testing Set)



Polynomial Fit (Testing Set)

From the above plots we can see that the polynomial fit isn't always better, as we see a decrease in prediction accuracy when going from linear to polynomial prediction with our Testing Set.

Appendix 1:
Problem 1-4 Steps:

$$\min_{\vec{\omega}, b} L(\vec{\omega}, b, \vec{a})$$

$$\min_{\vec{\omega}, b} \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j a_j \left[(\vec{\omega} \cdot \vec{x}^{(j)} + b) y_j - 1 \right]$$

$$= \min_{\vec{\omega}, b} \frac{1}{2} \sum_j \omega_j^2 - \overline{\qquad\qquad\qquad\qquad}$$

Minimize $\vec{\omega}$ by taking $\frac{\partial}{\partial \vec{\omega}} \left(L(\vec{\omega}, b, \vec{a}) \right)$

$$\frac{\partial}{\partial \vec{\omega}} \left[L(\vec{\omega}, b, \vec{a}) \right] = \vec{\omega} - \sum_j a_j \vec{x}^{(j)} y_j = 0$$

$$\Longrightarrow \vec{\omega} = \sum_j a_j y_j \vec{x}^{(j)}$$

$$\frac{\partial}{\partial b} \left[L(\vec{\omega}, b, \vec{a}) \right]$$

$$= - \sum_j a_j y_j = 0$$

$$\Longrightarrow 0 = \sum_j a_j y_j$$

Problem 1-5 Steps:

$$0 = \sum_j a_j y_j \quad (1)$$

$$\vec{\omega} = \sum_j a_j y_j \vec{x}^{(j)} \quad (2)$$

$$\min_{\vec{\omega}, b} \left(L(\vec{\omega}, b, \vec{a}) \right) = \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j a_j \left[(\vec{\omega} \cdot \vec{x}^{(j)} + b) y_j - 1 \right]$$

$$= \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j a_j \left[\vec{\omega} \cdot \vec{x}^{(j)} y_j + b y_j - 1 \right]$$

$$= \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j \left(a_j \vec{\omega} \cdot \vec{x}^{(j)} y_j + a_j b y_j - a_j \right)$$

$$= \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j \left(a_j \vec{\omega} \cdot \vec{x}^{(j)} y_j \right) - \sum_j (a_j b y_j) + \sum_j a_j$$

$$= \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j \left(a_j \vec{\omega} \cdot \vec{x}^{(j)} y_j \right) - b \cancel{\sum_j (a_j y_j)}^{\text{0 from (1)}} + \sum_j a_j$$

$$= \frac{1}{2} \vec{\omega} \cdot \vec{\omega} - \sum_j \left(a_j (\vec{\omega} \cdot \vec{x}^{(j)}) y_j \right) + \sum_j a_j$$

Plugging in (2)

$$= \underbrace{\frac{1}{2} \vec{\omega} \cdot \vec{\omega}}_{\text{looking closer}} - \sum_j \underbrace{\left(\left(\sum_k (a_k y_k \vec{x}^{(k)}) \cdot \vec{x}^{(j)} \right) a_j y_j \right)}_{\text{The same}} + \sum_j a_j$$

$$\frac{1}{2} \vec{\omega} \cdot \vec{\omega} = \frac{1}{2} \sum_j (a_j y_j \vec{x}^{(j)}) \cdot \sum_k (a_k y_k \vec{x}^{(k)})$$

$$\Rightarrow \max_{\vec{a}} \sum_j a_j - \frac{1}{2} \sum_j \sum_k y_j y_k a_j a_k (\vec{x}^{(j)} \cdot \vec{x}^{(k)})$$

Problem 1-8 Steps:

1-8

$$\max_{\vec{a}} \sum_{j} a_j - \frac{1}{2} \sum_{j} \sum_{k} y_j y_k a_j a_k \left(\vec{x}^{(j)} \cdot \vec{x}^{(k)}\right)$$

$\max_{\vec{a}}$ of following

$$a_1 + a_3 + a_4 - \frac{1}{2} \sum_{j} y_j y_1 a_j a_1 \left(\vec{x}^{(j)} \cdot \vec{x}^{(1)}\right)$$
$$- \frac{1}{2} \sum_{j} y_j y_3 a_j a_3 \left(\vec{x}^{(j)} \cdot \vec{x}^{(3)}\right)$$
$$- \frac{1}{2} \sum_{j} y_j y_4 a_j a_4 \left(\vec{x}^{(j)} \cdot \vec{x}^{(4)}\right)$$

$$= a_1 + a_3 + a_4 - \frac{1}{2} \Big[y_1 y_1 a_1 a_1 \left(\vec{x}^{(1)} \cdot \vec{x}^{(1)}\right) + y_3 y_1 a_3 a_1 \left(\vec{x}^{(3)} \cdot \vec{x}^{(1)}\right) + y_4 y_1 a_4 a_1 \left(\vec{x}^{(4)} \cdot \vec{x}^{(1)}\right) \Big]$$

$$-\frac{1}{2}\left[a_1^2 (2) + (-1)\, a_3 a_1 (1) + (-1)\, a_4 a_1 (1)\right] = -\frac{1}{2}\left[2a_1^2 - (a_3 + a_4) a_1\right]$$

$$- \frac{1}{2} \Big[y_1 y_3 a_1 a_3 \left(\vec{x}^{(1)} \cdot \vec{x}^{(3)}\right) + y_3 y_3 a_3 a_3 \left(\vec{x}^{(3)} \cdot \vec{x}^{(3)}\right) + y_4 y_3 a_4 a_3 \left(\vec{x}^{(4)} \cdot \vec{x}^{(3)}\right) \Big]$$

$$-\frac{1}{2}\left[(-1)\, a_1 a_3 (1) + (1)\, a_3^2 (1) + 0\right] = -\frac{1}{2}\left[a_3^2 - a_1 a_3\right]$$

$$- \frac{1}{2} \Big[y_1 y_4 a_1 a_4 \left(\vec{x}^{(1)} \cdot \vec{x}^{(4)}\right) + y_3 y_4 a_3 a_4 \left(\vec{x}^{(3)} \cdot \vec{x}^{(4)}\right) + y_4 y_4 a_4 a_4 \left(\vec{x}^{(4)} \cdot \vec{x}^{(4)}\right) \Big]$$

$$= -\frac{1}{2}\left[(-1)\, a_1 a_4 (1) + 0 + (1)\, a_4^2 (1)\right] = -\frac{1}{2}\left[a_4^2 - a_1 a_4\right]$$

$$\max_{\vec{a}} \Big(a_1 + a_3 + a_4 - \frac{1}{2}\left[2a_1^2 - (a_3+a_4)a_1\right] - \frac{1}{2}\left[a_3^2 - a_1 a_3\right] - \frac{1}{2}\left[a_4^2 - a_1 a_4\right] \Big)$$

$$a_1 + a_3 + a_4 - a_1^2 + \frac{a_1 a_3}{2} + \frac{a_1 a_4}{2} - \frac{a_3^2}{2} + \frac{a_1 a_3}{2} - \frac{a_4^2}{2} + \frac{a_1 a_4}{2}$$

$$= a_1 + a_3 + a_4 - a_1^2 + a_1 a_3 + a_1 a_4 - \frac{a_3^2}{2} - \frac{a_4^2}{2}$$

$a_1 = a_3 + a_4$

$$\Rightarrow a_3 + a_4 + a_3 + a_4 - (a_3+a_4)^2 + (a_3+a_4)a_3 + (a_3+a_4)a_4 - \frac{a_3^2}{2} - \frac{a_4^2}{2}$$

$$= 2a_3 + 2a_4 - \cancel{a_3^2} - \cancel{a_4^2} - \cancel{2a_3 a_4} + \cancel{a_3^2} + \cancel{a_3 a_4} + \cancel{a_3 a_4} + \cancel{a_4^2} - \frac{a_3^2}{2} - \frac{a_4^2}{2}$$

$$= 2a_3 + 2a_4 - \frac{a_3^2}{2} - \frac{a_4^2}{2}$$

$$\max_{\vec{a}} \left(2a_3 + 2a_4 - \frac{a_3^2}{2} - \frac{a_4^2}{2} \right)$$

Take partial derivative wrt a_3 & a_4
set to zero & solve.

$$\frac{\partial}{\partial a_3}(\dots) = 2 - a_3 = 0$$

$$\Rightarrow a_3 = a_4 = 2$$

$$\frac{\partial}{\partial a_4}(\dots) = 2 - a_4 = 0$$

$$\Rightarrow a_1 = a_3 + a_4 = 4$$

Problem 1-9 Steps:

1- 9

$$\vec{\omega} = \sum_j a_j y_j \vec{x}^{(j)}$$

$$= a_1 y_1 \vec{x}^{(1)} + a_3 y_3 \vec{x}^{(3)} + a_4 y_4 \vec{x}^{(4)}$$

$$= 4(1)\begin{bmatrix} 1 \\ 1 \end{bmatrix} + 2(-1)\begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2(-1)\begin{bmatrix} G \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & -2 \\ 4 & -2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\vec{\omega} = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$