Name: Yihan Zhou **GT account:** yzhou376@gatech.edu
GT number: 903053761

CS 4641 Machine Learning
HW2

Problems Set

1. To set α_k to be constant means the learning rate will not change throughout the whole learning process. While to set α_k as function k means the learning rate is adaptable to learning time.

2.a. Because $\mathbf{w}$ is perpendicular to the decision boundary, it should be parallel to $\phi(x_2) - \phi(x_1)$. which is $\begin{bmatrix} 1 \\ \sqrt{2}\cdot\sqrt{2} \\ \sqrt{2}^2 \end{bmatrix} - \begin{bmatrix} 1 \\ \sqrt{2}\cdot 0 \\ 0^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$

2.b. Margin is the Euclidean distance between the two points. We get $\sqrt{0^2+2^2+2^2} = 2\sqrt{2}$.

2.c. According to definition, $\sqrt{2} = \frac{2}{\|\mathbf{w}\|_2}$. Thus we solve $\|\mathbf{w}\|_2 = \frac{1}{\sqrt{2}}$. Since $\mathbf{w}$ is parallel to $\begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$, we get $\mathbf{w} = \begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix}$.

2.d. We substitute the value of $\mathbf{w}, y_1$, and y_2 back into the inequality and get

$$-1 \cdot (\begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{2}\cdot 0 \\ 0^2 \end{bmatrix} + w_0) \geq 1$$

$$1 \cdot (\begin{bmatrix} 0 \\ 0.5 \\ 0.5 \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{2}\cdot\sqrt{2} \\ \sqrt{2}^2 \end{bmatrix} + w_0) \geq 1$$

By calculation, we get

$$-w_0 \geq 1$$

$$2 + w_0 \geq 1$$

$$w_0 \leq -1$$

Thus,

$$w_0 = -1$$

.

2.e. We substitute all the elements back and get

$$h(x) = -1 + \frac{\sqrt{2}}{x}x + \frac{1}{2}x^2$$

Implementation Short Answers

2.3. As we increase the lambda, we blur the boundary, which leads probability to be less deterministic. As we decrease the lambda, the result turned to be more polarized. As we increase the lambda, bias increase and variance decrease. As we decrease the lambda, bias decrease and variance increase. (Plots are attached at the end of the pdf. Please check them there. Thx!!)

3.4. Poly: As we increase the degree in SVM polynomial Kernel, the decision boundary tends to be increasingly complicate curve. Biases decreases while variances increases, which cause over-fitting on training data. As we increase the C in SVM, the decision boundary also increasingly classifies all training data correct, but tends to lose natural separation between the data.Biases decreases while variances increases, which also cause over-fitting on training data.

Gaussian: As we decrease the sigma in SVM Gaussian Kernel, the decision boundary tends to be increasingly fits every data, and finally converge around some data point. Biases decreases while variances increases, which cause over-fitting on training data. As we increase the C in SVM, the decision boundary also increasingly classifies all training data correct, decreases biases while increases variances,causing over-fitting on training data.

Figure 1: Lambda=0.0000001(Probability)

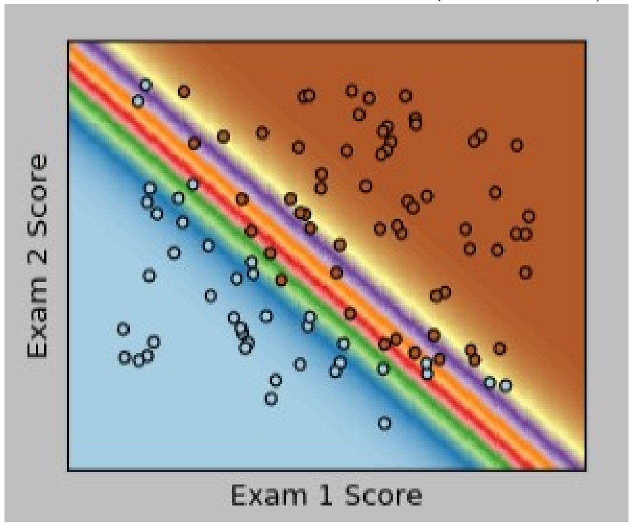

Figure 2: Lambda=0.0000001(Threshold)

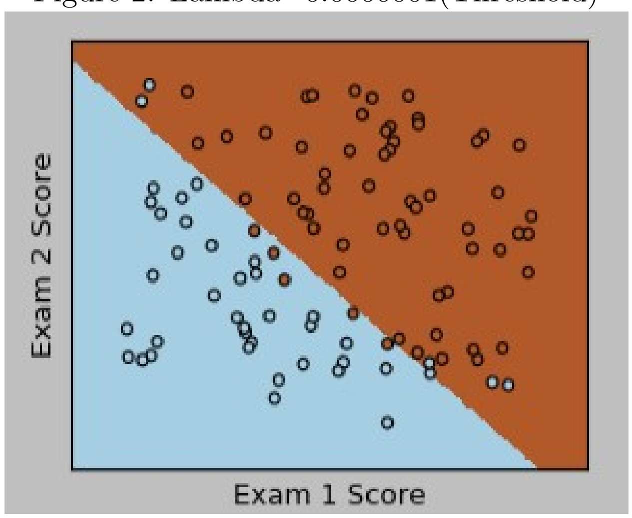

Figure 3: Lambda=10(Probability)

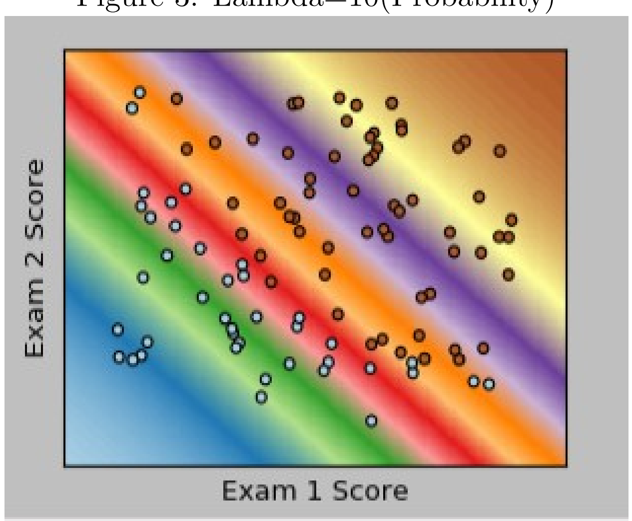

Figure 4: Lambda=10(Threshold)

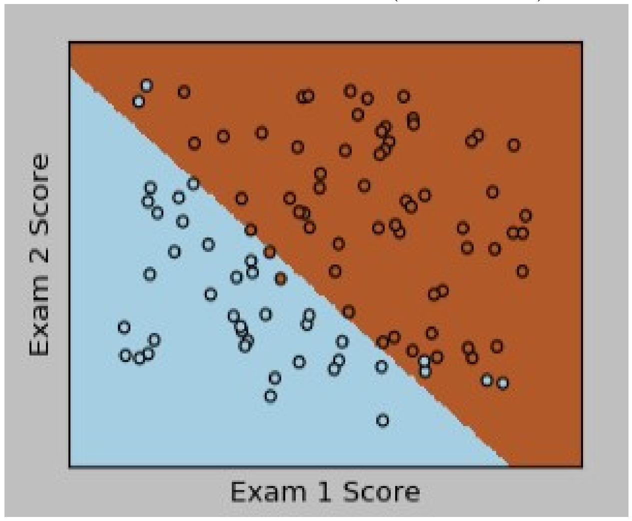

Figure 5: Lambda=100(Probability)

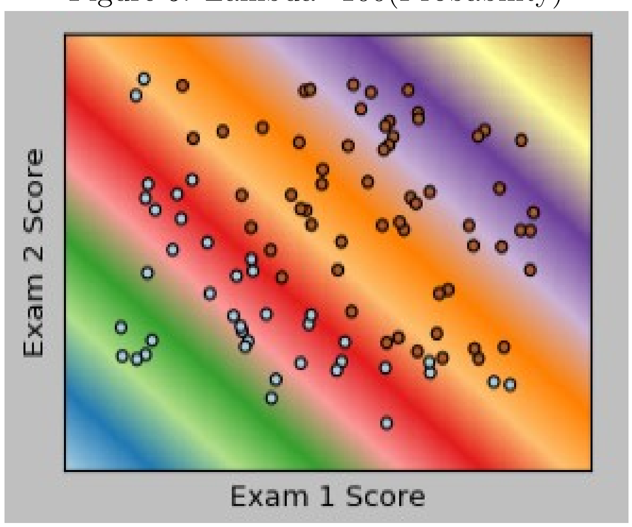

Figure 6: Lambda=100(Threshold)