

CIS 419 – Homework 2

Devin Stein
devstein@seas.upenn.edu

October 15, 2016

Problem 1

Using a constant α_k convergence will take longer because if α_k is too small it will move slowly towards convergence and if α_k is too large it will overstep the convergence point.

However, using a α_k as a function of k the step size will change as it approaches the converges point reaching the point faster and more accurately.

Problem 2

a) Write down a vector that is parallel to the optimal vector w .

$$\phi_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \phi_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \quad \phi_2 - \phi_1 = \begin{bmatrix} 0 \\ 2 \\ 2 \end{bmatrix}$$

b) What is the value of margin achieved by w ?

$$\|w\|_2 = \sqrt{8} = 2\sqrt{2}$$

c) Solve for w .

$$\frac{2}{\|w\|_2} = \frac{2}{\sqrt{8}} = \frac{1}{\sqrt{2}} = w$$

d) Solve for w_0 .

$$\begin{bmatrix} 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + w_0 \geq 1$$

$$\begin{bmatrix} 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} + w_0 \geq 1$$

$$w_0 \geq 1$$

$$w_0 = 1$$

e) Write down the discriminant form $h(x) = w_0 + w^T \phi(x)$.

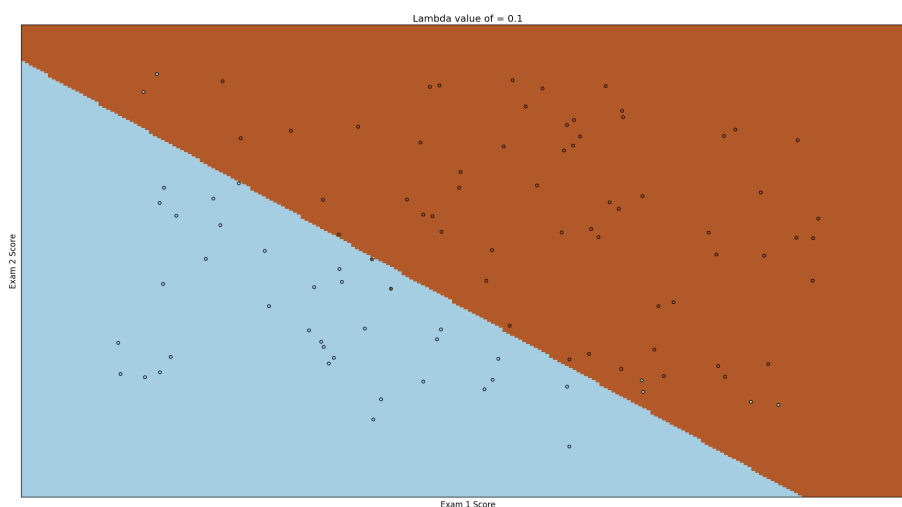
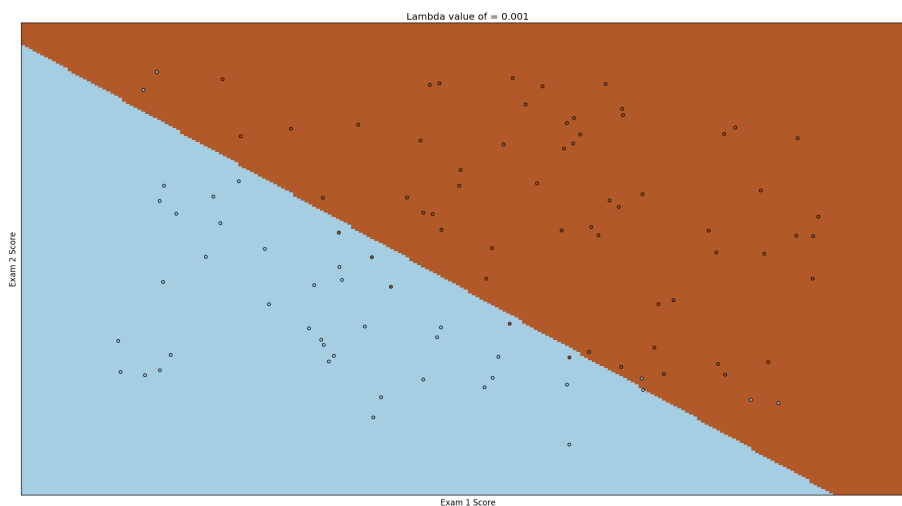
$$h(x) = 1 + \begin{bmatrix} 0 & 2 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ \sqrt{2}x \\ x^2 \end{bmatrix}$$

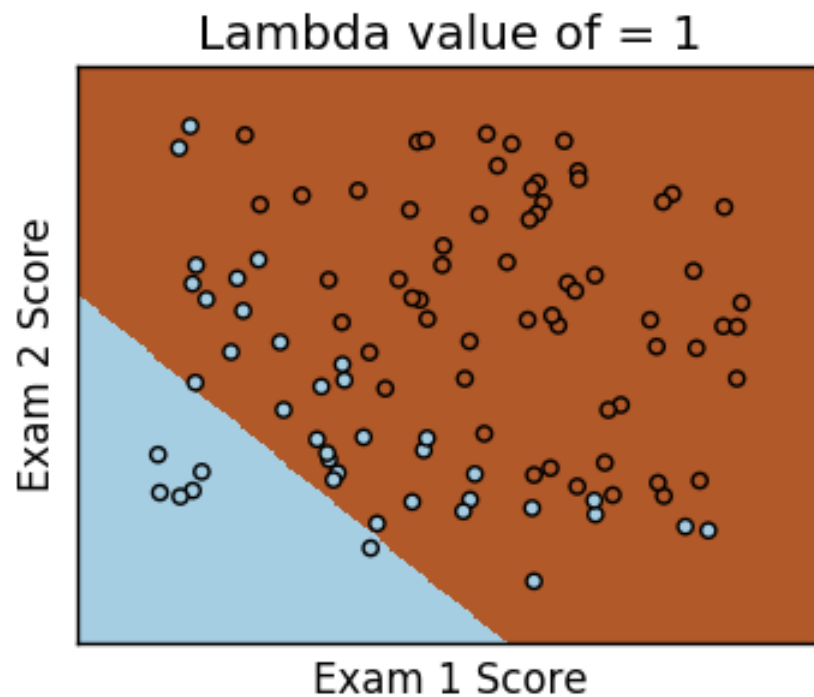
$$h(x) = 1 + 2\sqrt{2}x + 2x^2$$

Draw the complete (unpruned) decision tree, showing the class predictions at the leaves.

Problem 3

As I increased the value of lambda, I observed that the predicted line was shifted in the south west direction. Capturing more of the blue dots in the red region as the value increased. At a value of 1 about 3/4 of the figure was shaded red. The increase in lambda causes the gradient to increase more rapidly causing it to converge faster. See pictures below.





Problem 4

As C increased, I observed that the algorithm overclassifies capturing as many points as possible even though they are extreme outliers. As σ increases, the boundaries of the classification become more rigid with straight lined edges thus reducing the variances of the gaussian. As d increases in the polynomial kernel, the polynomial curve of the boundaries increase accordingly. See pictures below.

