

# **CIS 520: Problem Set #4**

Due on October 15, 2017

*Lyle Ungar, Shivani Agarwal*

**Eric Oh**

Collaborators: Jiarui Lu

## Problem 1

Convolutional Neural Network

### Solution

1. Each neuron in the first hidden layer would require 48150 weights for a fully connected neural network.
2. Using a filter of size  $21 \times 14 \times 3$  results in 882 weights for a single neuron.
3. The size of the output image from the Conv layer is

$$\frac{W - F + 2P}{S} + 1$$

where  $W$  is the input volume size,  $F$  is the size of the Conv layer filter,  $P$  is the amount of zero padding on the border, and  $S$  is the stride size. Thus, for a  $105 \times 154$  input image, a  $21 \times 14$  filter, 0 padding, and a stride of 7, the output image is  $13 \times 21$ , resulting in 273 neurons.

4. The first output of the  $3 \times 3 \times 2$  output is given by applying the  $3 \times 3 \times 1$  filter 1 to the input image with a stride of 1. The second is found by applying the  $3 \times 3 \times 1$  filter 2 to the input image with a stride of 1. This results in an output of

$$\begin{pmatrix} 0 & -2 & 0 \\ -2 & -4 & -2 \\ 0 & -2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 3 & 0 \\ 3 & 6 & 3 \\ 0 & 3 & 0 \end{pmatrix}$$

## Problem 2

Convex Sets

### Solution

1. (a) Yes, this is a convex set because it is a Euclidean norm ball in  $\mathbb{R}^3$  centered at the origin with radius  $\sqrt{8}$ .  
 (b) No, this is not a convex set. Take the points  $(0, 0, 3)$  and  $(0, 0, -3)$  both in  $C$ . The line segment connecting the two points contains the origin, which is not in  $C$ .  
 (c) No, this is not a convex set. Take the points  $(0, 4, 0)$  and  $(0, 0, 4)$  both in  $C$ . The line segment connecting the two points contains  $(0, 2, 2)$ , which is not in  $C$ .  
 (d) Yes, this is a convex set. Note that each of the constraints can be written as the intersection of two halfspaces. For example,  $0 \leq x_1 \leq 1$  is equivalent to the intersection of  $a_1^T x + b_1 \leq 0$  and  $a_2^T x + b_2 \leq 0$  where  $a_1 = (-1, 0, 0)$ ,  $b_1 = 0$ ,  $a_2 = (1, 0, 0)$ ,  $b_2 = -1$ . Thus,  $C$  is the intersection of halfspaces and is thus a convex set.  
 (e) Yes, this is a convex set as it is the intersection of  $l_\infty$  and  $l_1$  norm balls.
2. (a)  $f$  is both convex and concave due to it being an affine function.  
 (b)  $f$  is not a convex function. To see this, note that we can write  $f$  as  $f(x) = x^T A x$  where  $A =$

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$f$  would be a convex function if  $A$  is positive semi-definite. Since  $A$  is a diagonal matrix, the eigenvalues are given by the diagonal entries and thus we check if either  $A$  or  $-A$  has all non-negative eigenvalues. We note that since neither  $A$  nor  $-A$  has all non-negative eigenvalues,  $A$  is not positive semi-definite and thus  $f$  is not convex.

(c)  $f$  is a convex set. To see this, note that we can write  $f$  as  $f(x) = x^T A x$  where  $A =$

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

$f$  would be a convex function if  $A$  is positive semi-definite. Since  $A$  is a diagonal matrix, the eigenvalues are given by the diagonal entries and thus we check if either  $A$  or  $-A$  has all non-negative eigenvalues. We note that since  $-A$  has all non-negative eigenvalues,  $A$  is positive semi-definite and thus  $f$  is convex.

(d)  $f$  is convex because it is the supremum of convex functions.

(e)  $f$  is convex because it is the  $L_1$  norm of  $x$ .

## Problem 3

Optimization and Duality

**Solution**

$$1. L(x_1, x_2, \lambda) = 4x_1 + 3x_2 + \lambda(x_1^2 + x_2^2 - 1)$$

2. The minimum values of  $x_1$  and  $x_2$  are given by

$$\begin{aligned} \frac{dL}{dx_1} = 4 + 2\lambda x_1 &\equiv 0 &\Rightarrow & \hat{x}_1 = -\frac{2}{\lambda} \\ \frac{dL}{dx_2} = 3 + 2\lambda x_2 &\equiv 0 &\Rightarrow & \hat{x}_2 = -\frac{3}{2\lambda} \end{aligned}$$

Note that the second derivatives are positive, indicating that they are indeed minima. The dual function is then given by

$$\phi(\lambda) = \inf_x L(x_1, x_2, \lambda) = 4 \left( -\frac{2}{\lambda} \right) + 3 \left( -\frac{3}{2\lambda} \right) + \lambda \left( \frac{4}{\lambda^2} \right) + \lambda \left( \frac{9}{4\lambda^2} \right) - \lambda = -\frac{4}{\lambda} - \frac{9}{4\lambda} - \lambda$$

3. To solve the dual problem, we initially attempt to find an unconstrained maximum of the dual objective. If this maximum is greater than 0, then we have a solution to the constrained dual problem.

$$\frac{d\phi(\lambda)}{d\lambda} = \frac{4}{\lambda^2} + \frac{9}{4\lambda^2} - 1 \equiv 0 \quad \Rightarrow \quad \hat{\lambda} = \pm \frac{5}{2}$$

Since  $\hat{\lambda} = \frac{5}{2}$  is greater than 0, we have a solution to the constrained dual problem. Thus, a solution to the primal problem is given by  $\hat{x}_1 = -\frac{2}{2.5} = -0.8$  and  $\hat{x}_2 = -\frac{3}{5} = -0.6$ . This solution does not lie in the Euclidean ball of the primal constraint because  $(-0.8)^2 + (-0.6)^2 = 1$ , rather it lies on it.

## Problem 4

Kernel Functions

**Solution**

1. Let  $\phi(x) = \sqrt{c}\phi_1(x)$ . Then

$$\phi(x)^T \phi(x') = \sqrt{c}\phi_1^T(x) \sqrt{c}\phi_1(x') = c\phi_1^T(x)\phi_1^T(x') = cK_1(x, x') = K(x, x')$$

2. Let  $\phi(x) = (\phi_1(x), \phi_2(x))$ . Then

$$\begin{aligned}\phi(x)^T \phi(x') &= (\phi_1(x), \phi_2(x))^T (\phi_1(x'), \phi_2(x')) \\ &= \phi_1^T(x) \phi_1(x') + \phi_2^T(x) \phi_2(x') \\ &= K_1(x, x') + K_2(x, x') = K(x, x')\end{aligned}$$

3. A mapping  $\phi$  cannot exist because dot products always sum elements. If either  $\phi_1(x)$  or  $\phi_2(x)$  were made to be negative, the dot product would then result in a summation.

4. Let  $\phi(x) = \phi_1(f(x))$ . Then

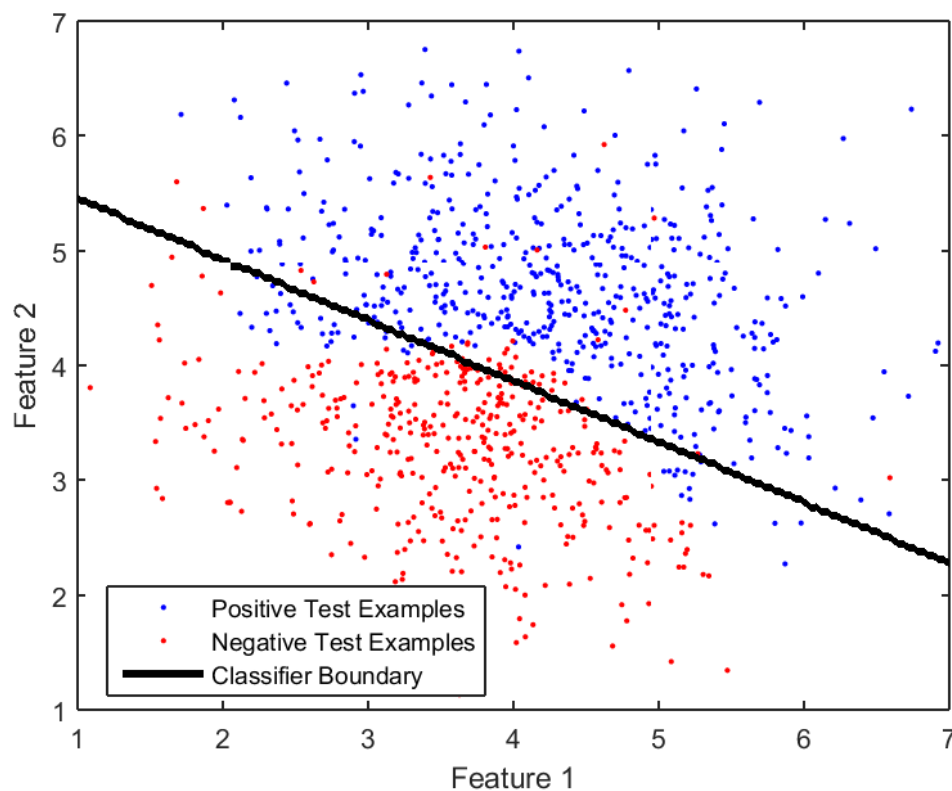
$$\phi(x)^T \phi(x') = \phi_1^T(f(x)) \phi_1(f(x')) = K_1(f(x), f(x')) = K(x, x')$$

## Problem 5

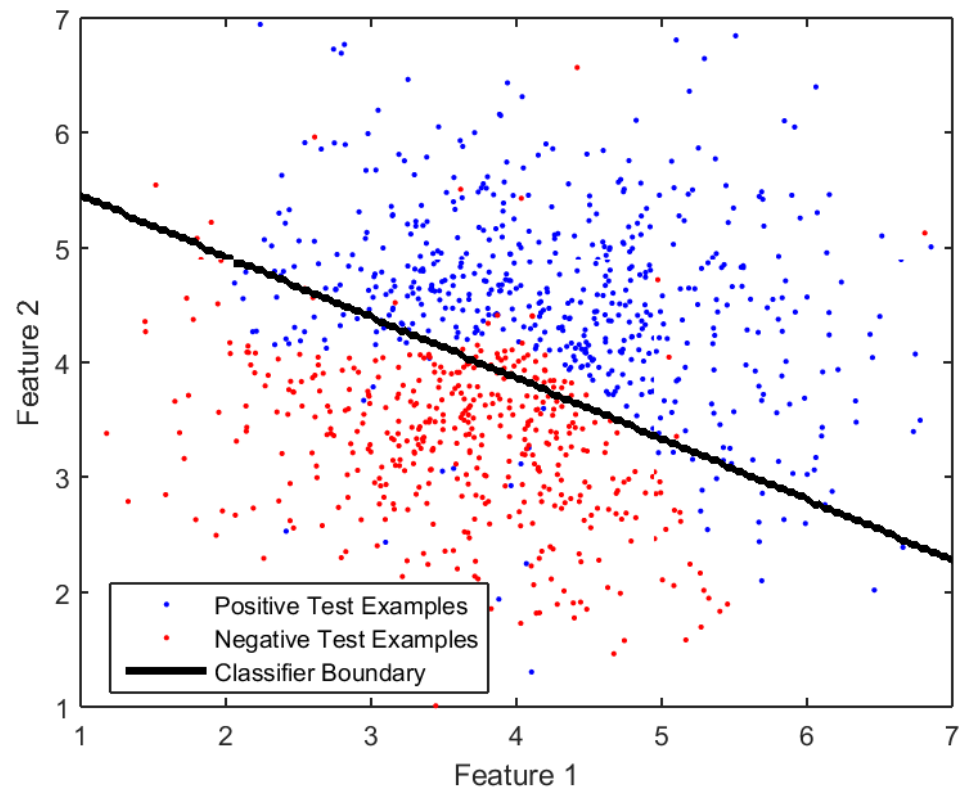
SVM and Kernels

### Solution

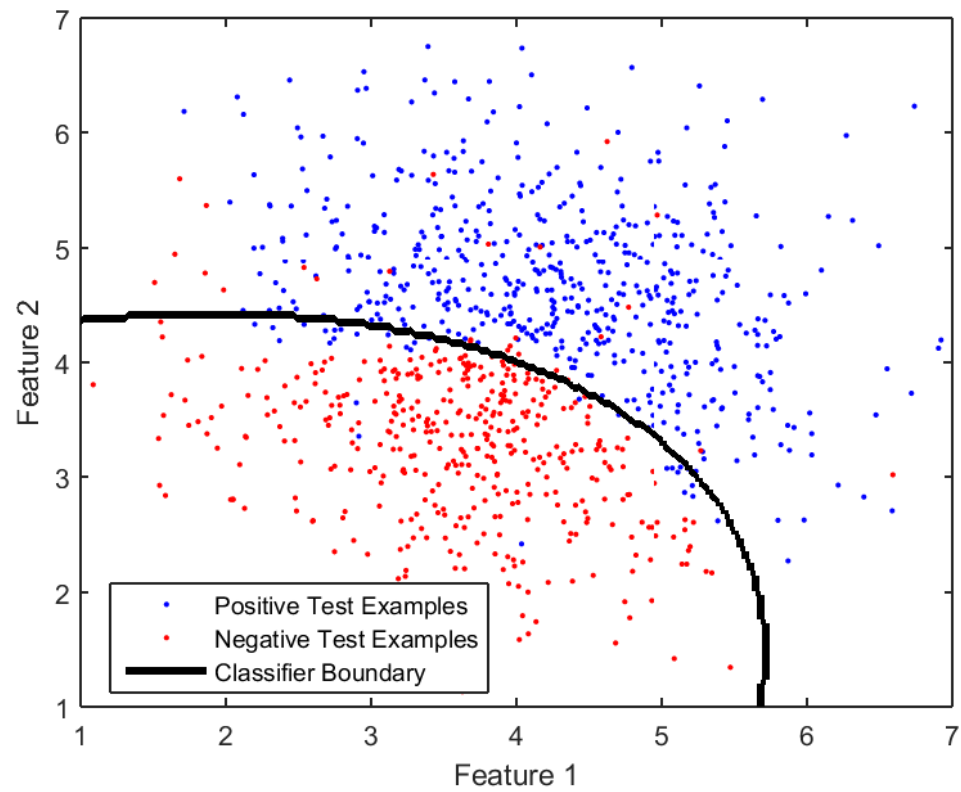
1. • For  $q = 1$ , the training decision boundary is shown in the plot below.



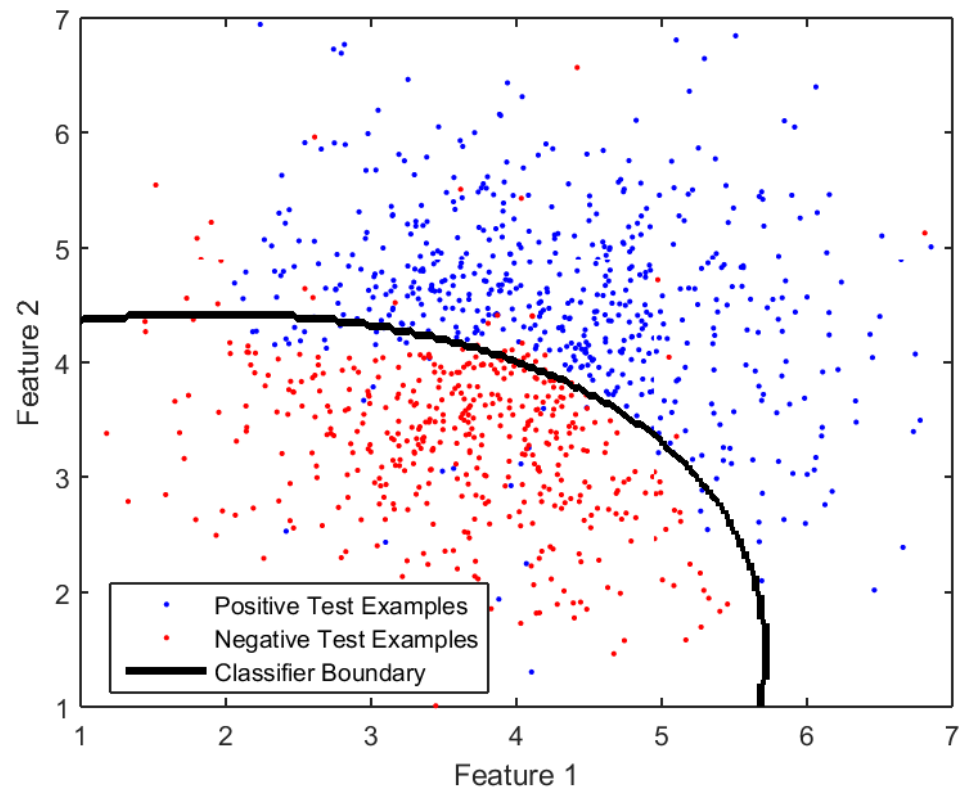
The testing decision boundary is shown below.



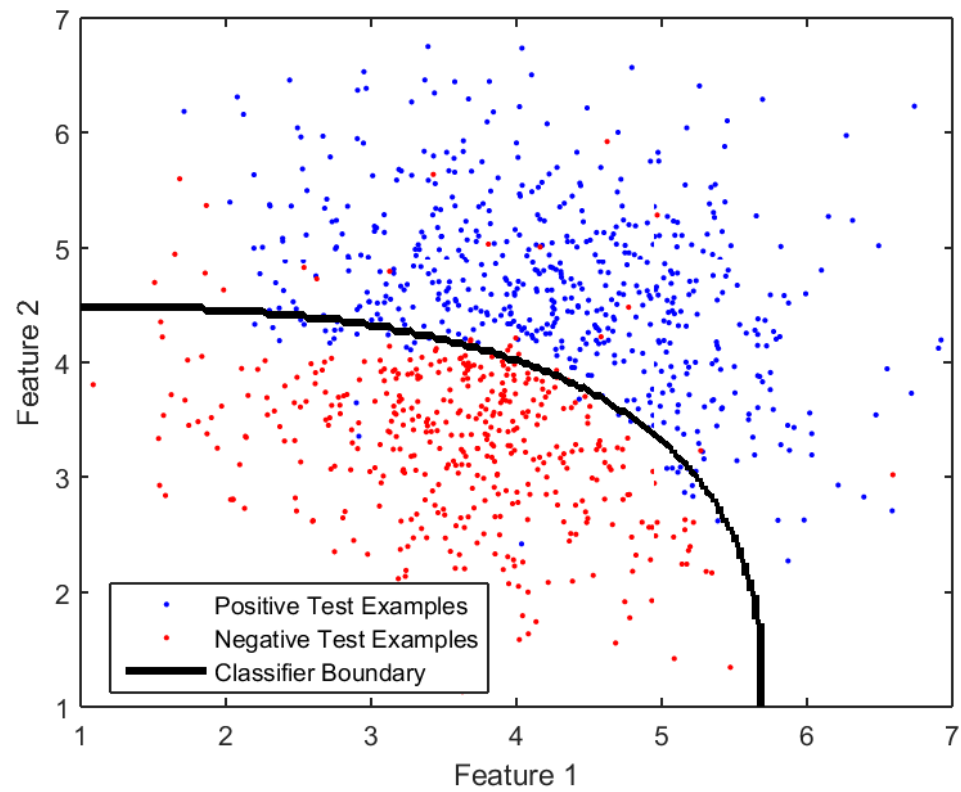
- For  $q = 2$ , the training decision boundary is shown in the plot below.



The testing decision boundary is shown below.

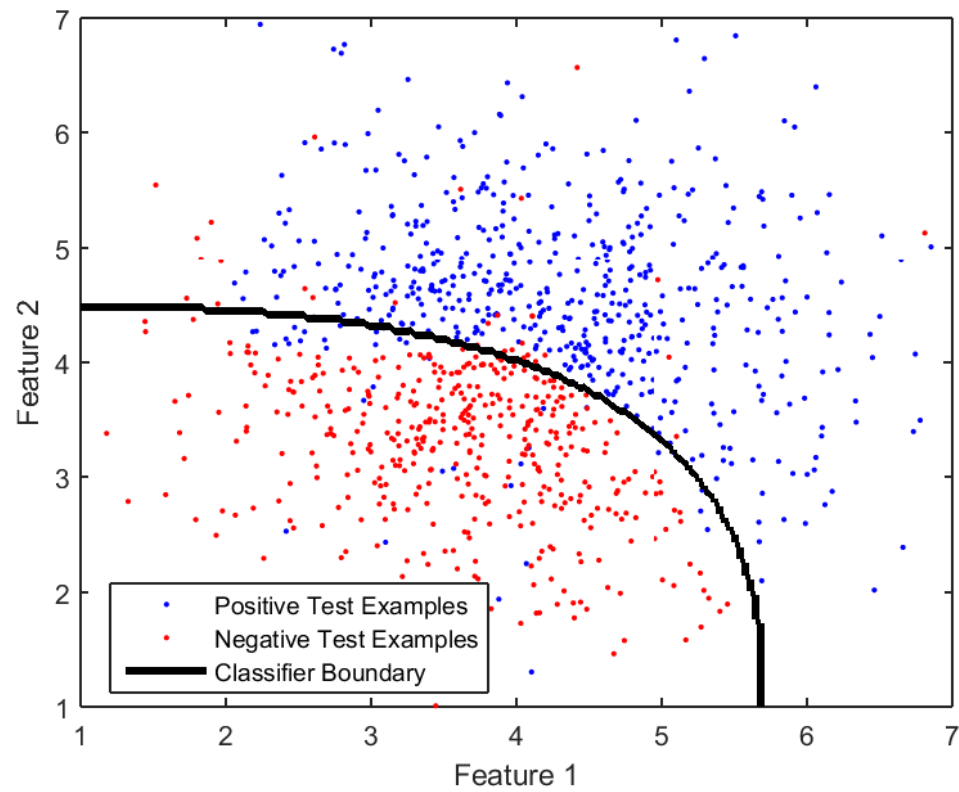


- For  $q = 3$ , the training decision boundary is shown in the plot below.

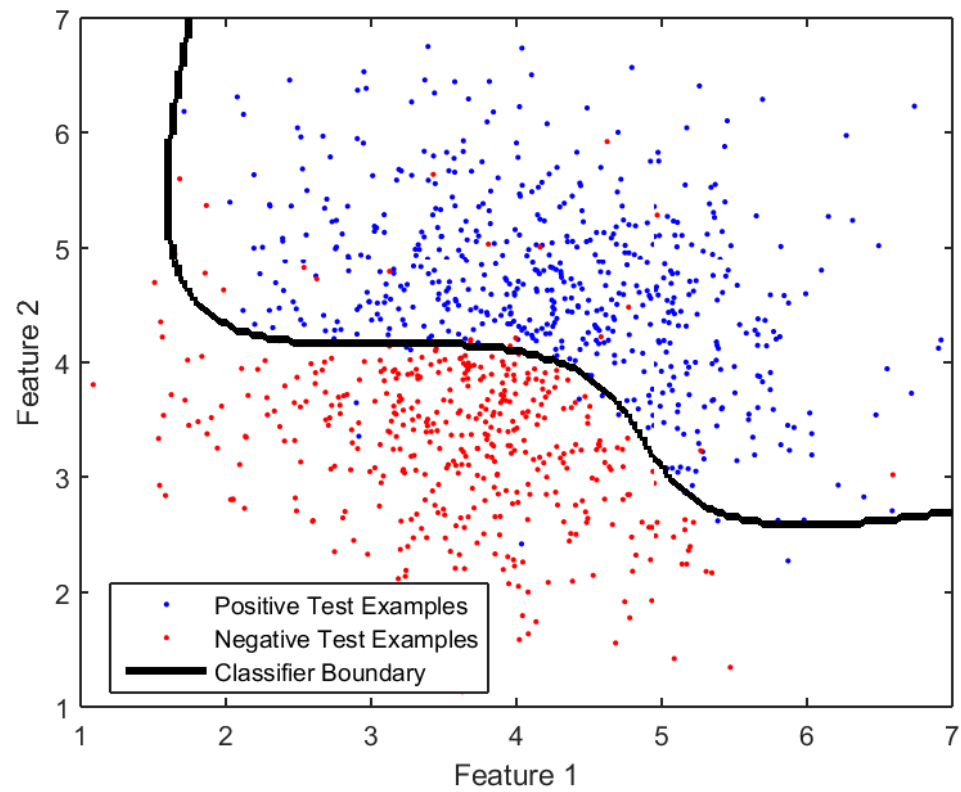


The testing decision boundary is shown below.

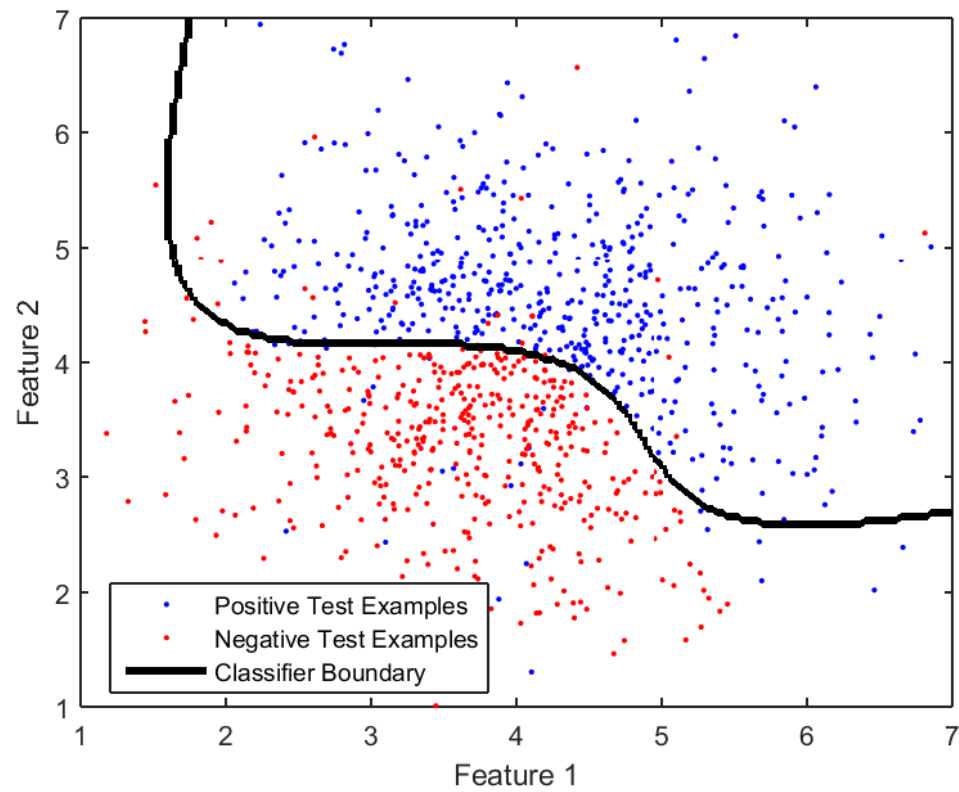




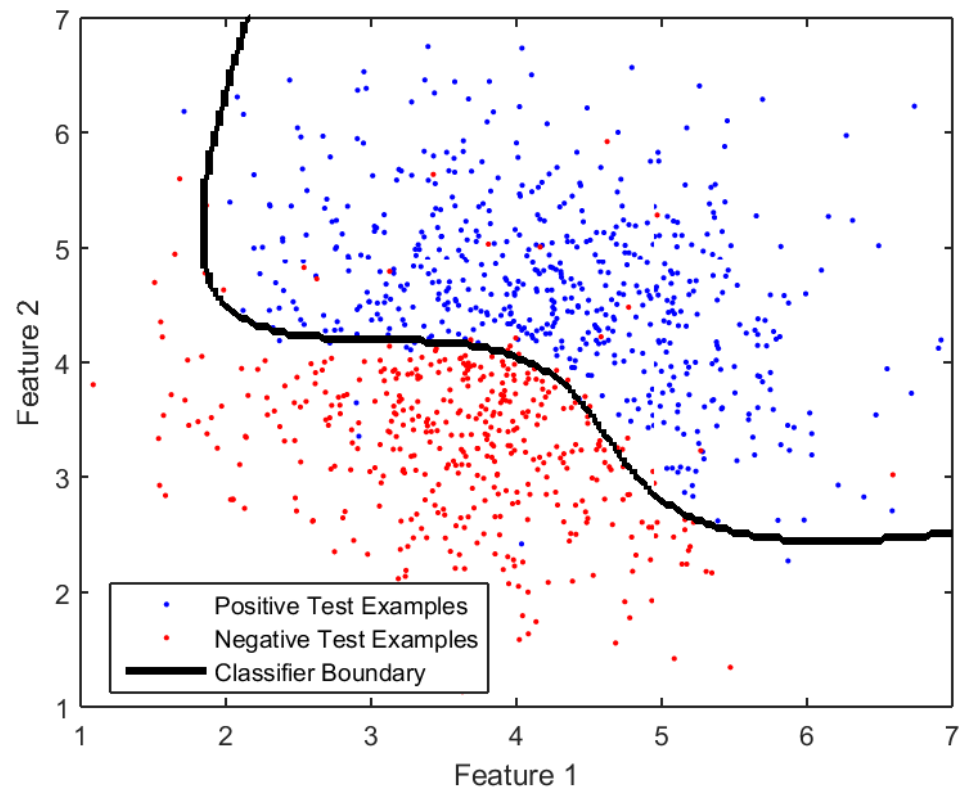
- For  $q = 4$ , the training decision boundary is shown in the plot below.



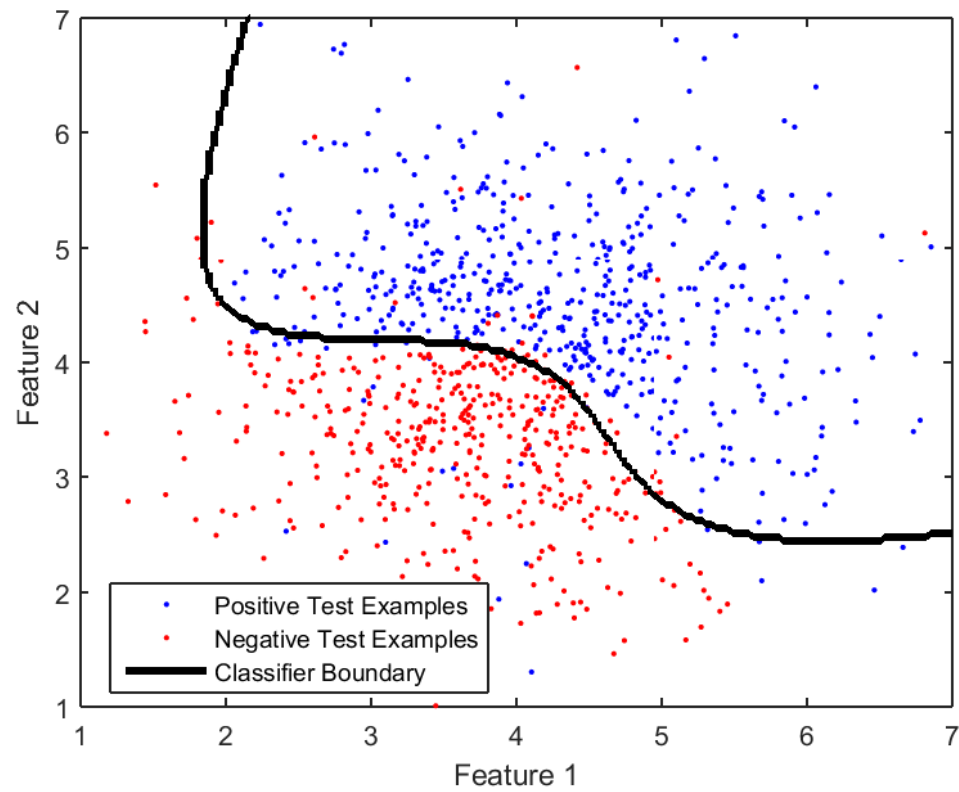
The testing decision boundary is shown below.



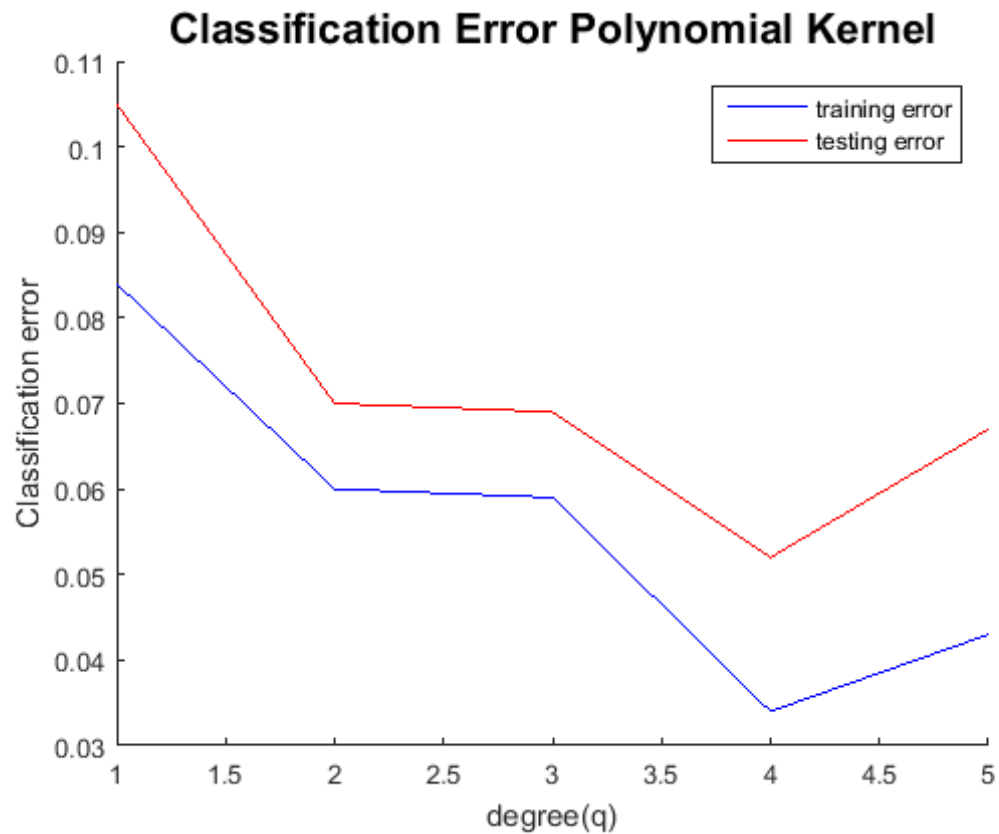
- For  $q = 5$ , the training decision boundary is shown in the plot below.



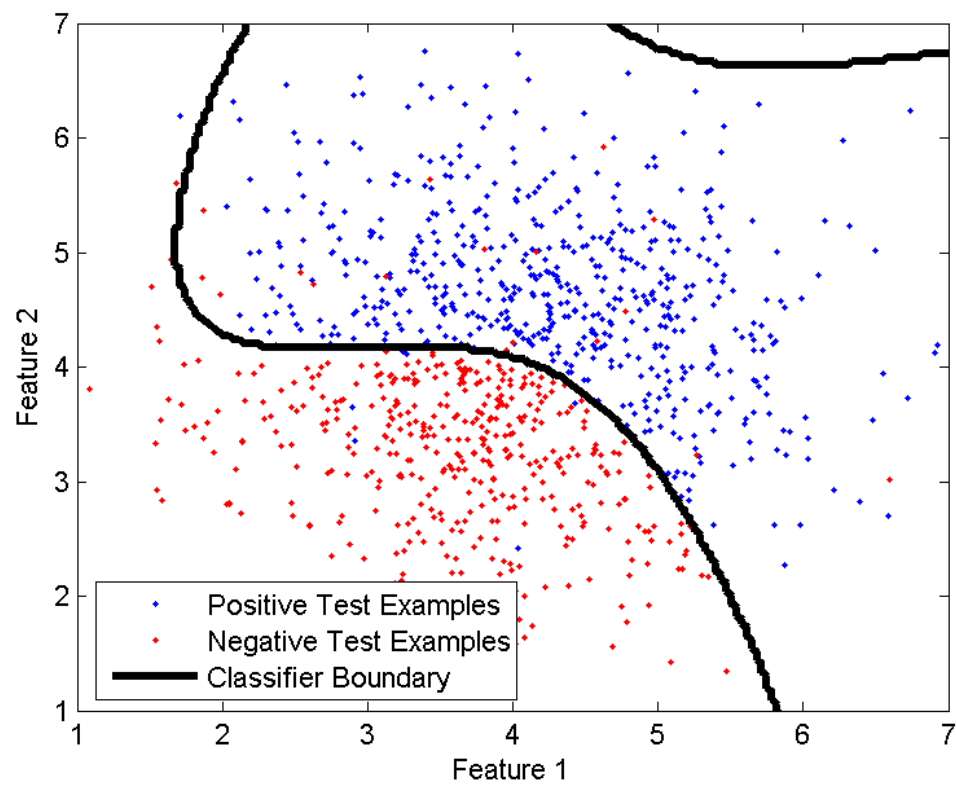
The testing decision boundary is shown below.



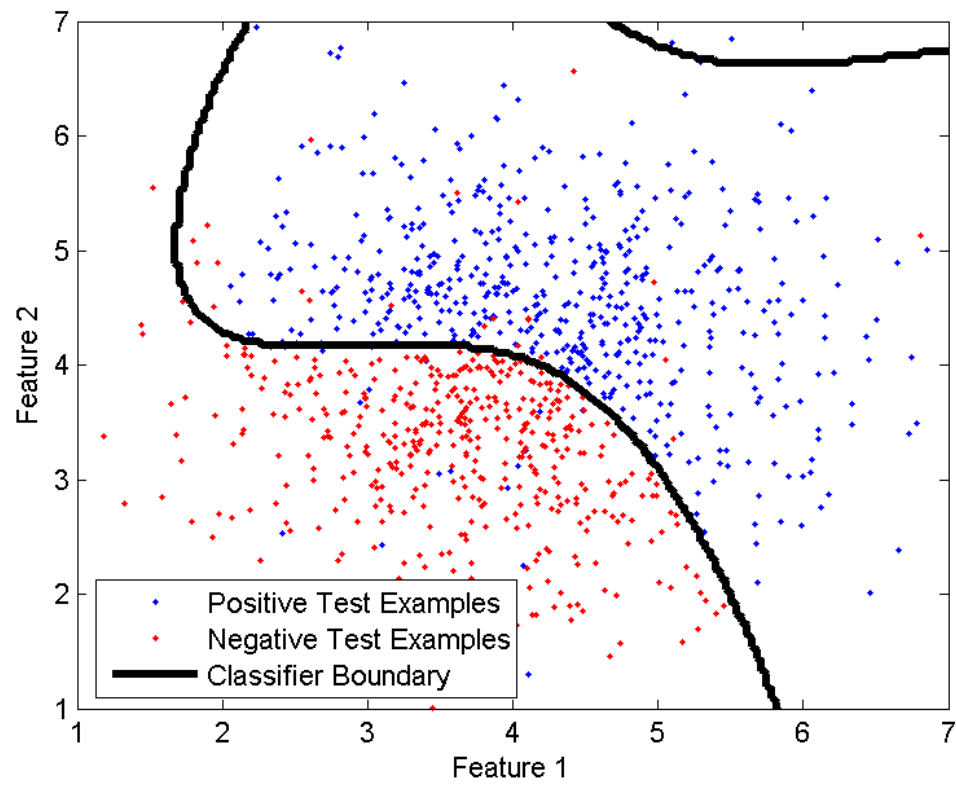
- The training and testing error achieved by each value of  $q$  is shown in the plot below. The value of  $q$  that achieves the lowest test error is  $q = 4$ .



2. • For  $\sigma = 0.01$ , the training decision boundary is shown in the plot below.

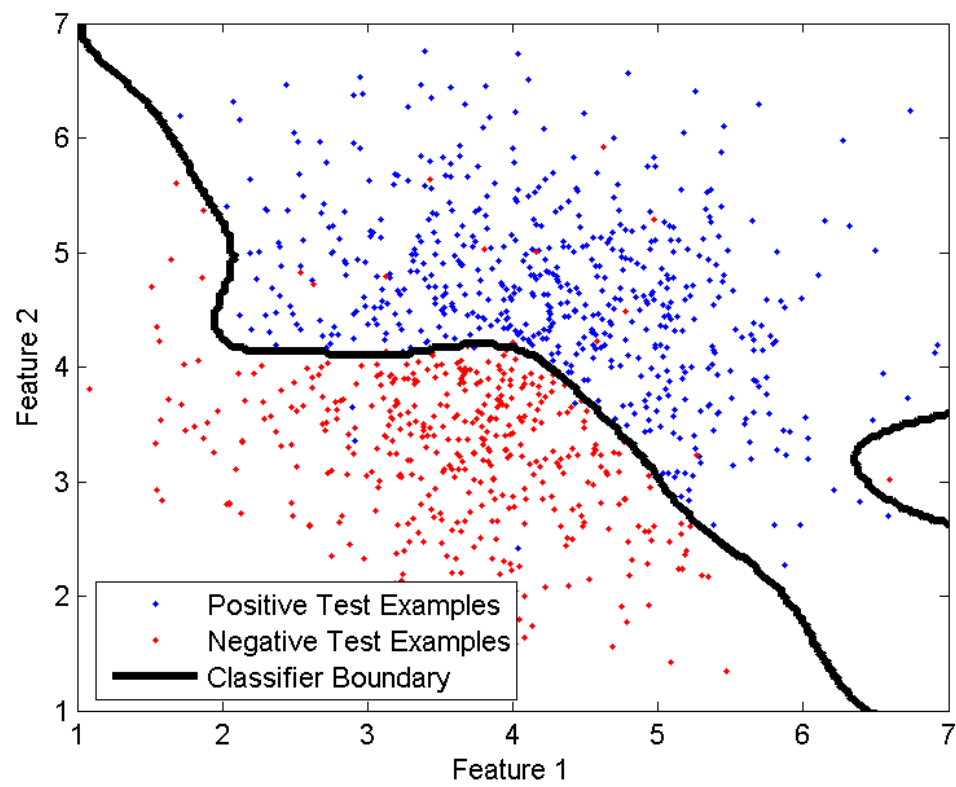


The testing decision boundary is shown below.

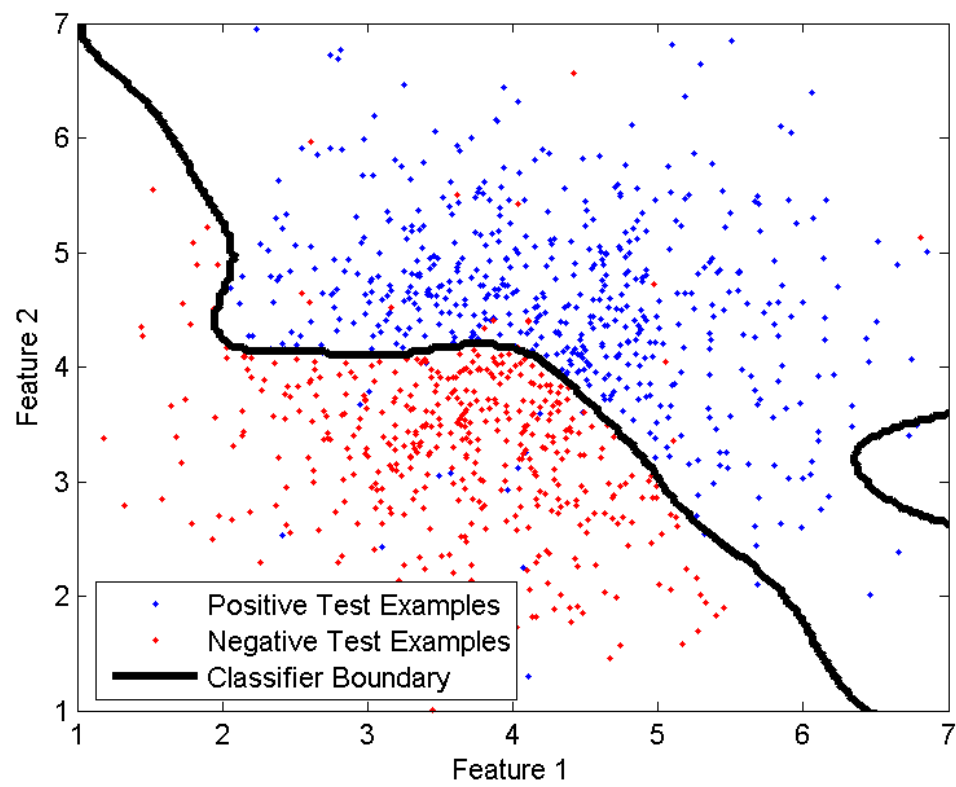


- For  $\sigma = 1$ , the training decision boundary is shown in the plot below.

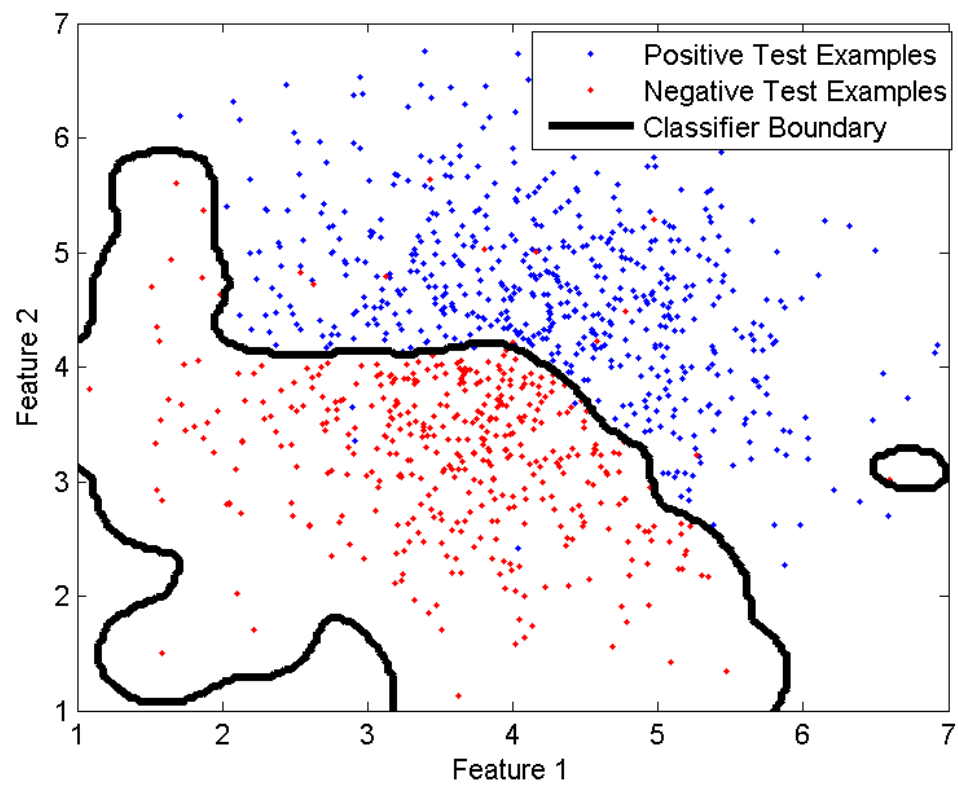




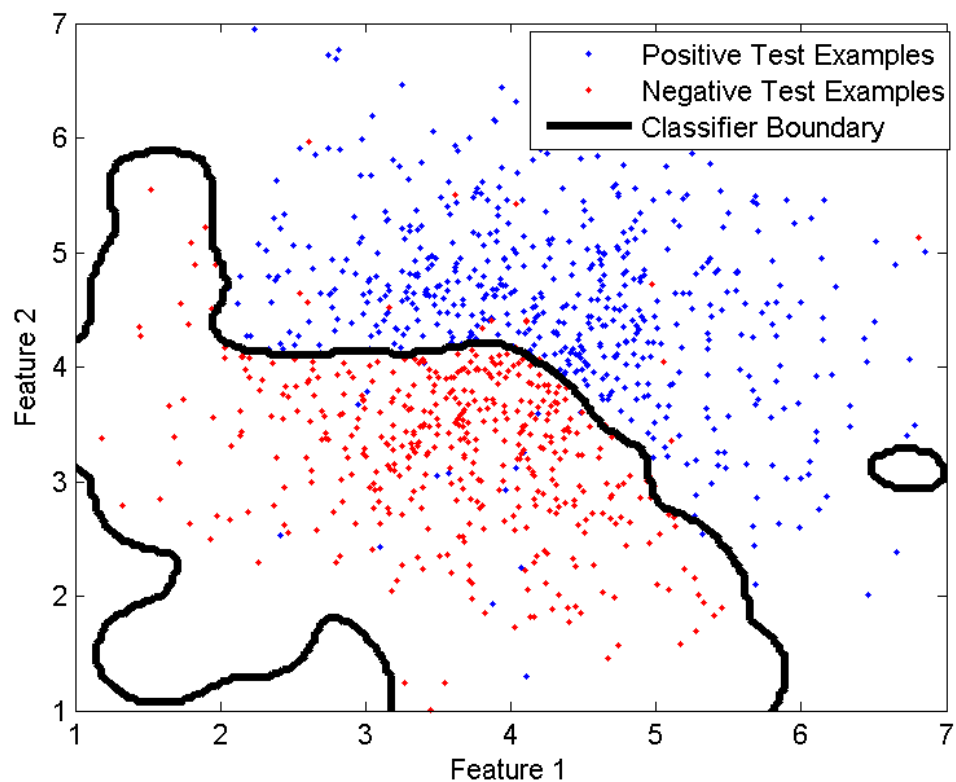
The testing decision boundary is shown below.



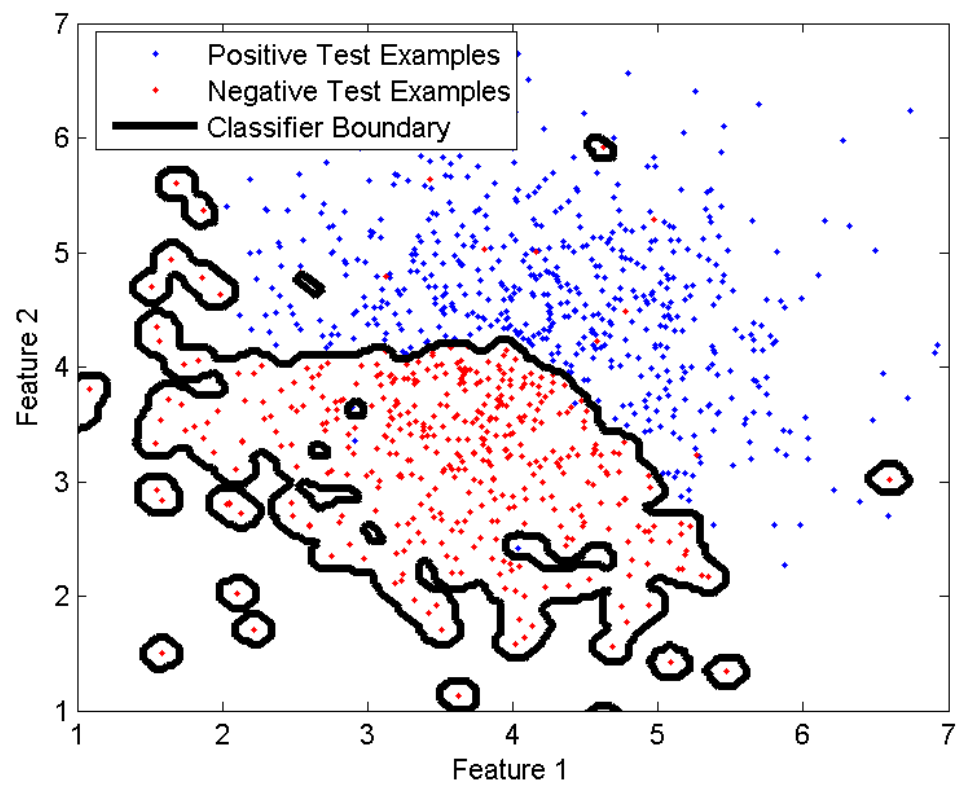
- For  $\sigma = 10$ , the training decision boundary is shown in the plot below.



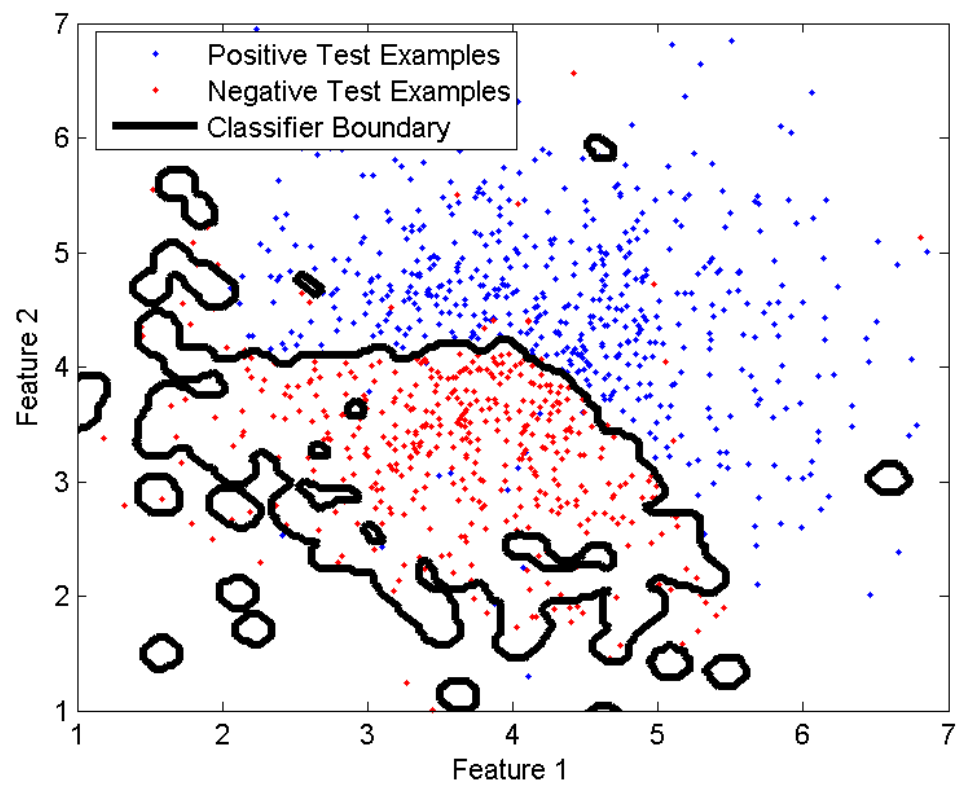
The testing decision boundary is shown below.



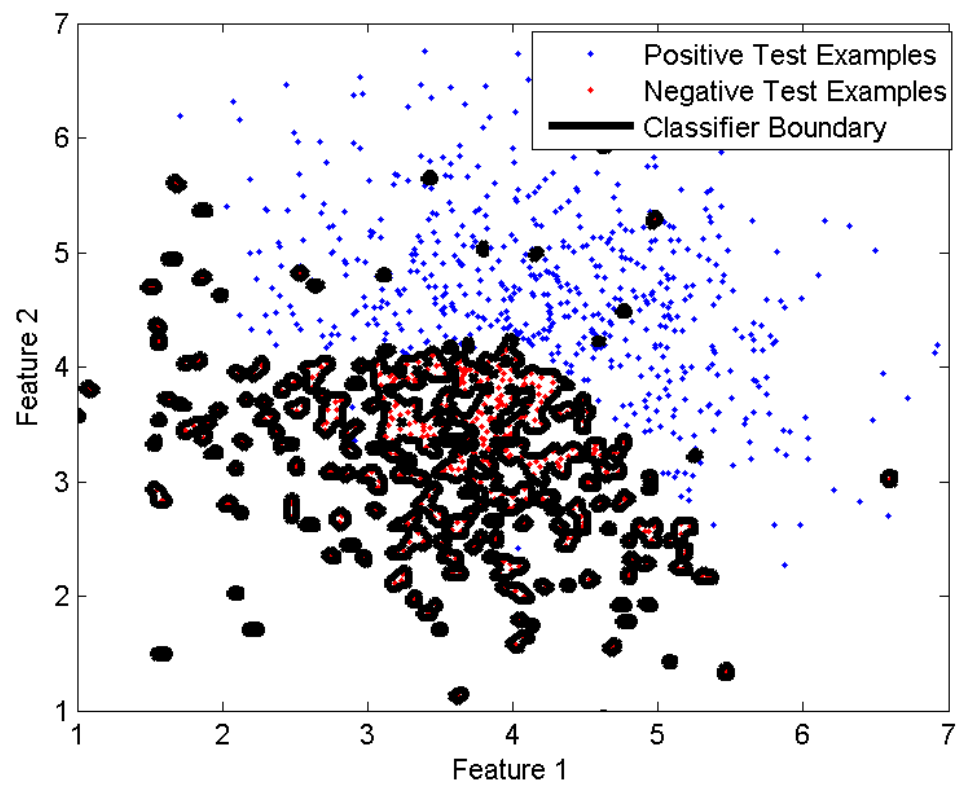
- For  $\sigma = 100$ , the training decision boundary is shown in the plot below.



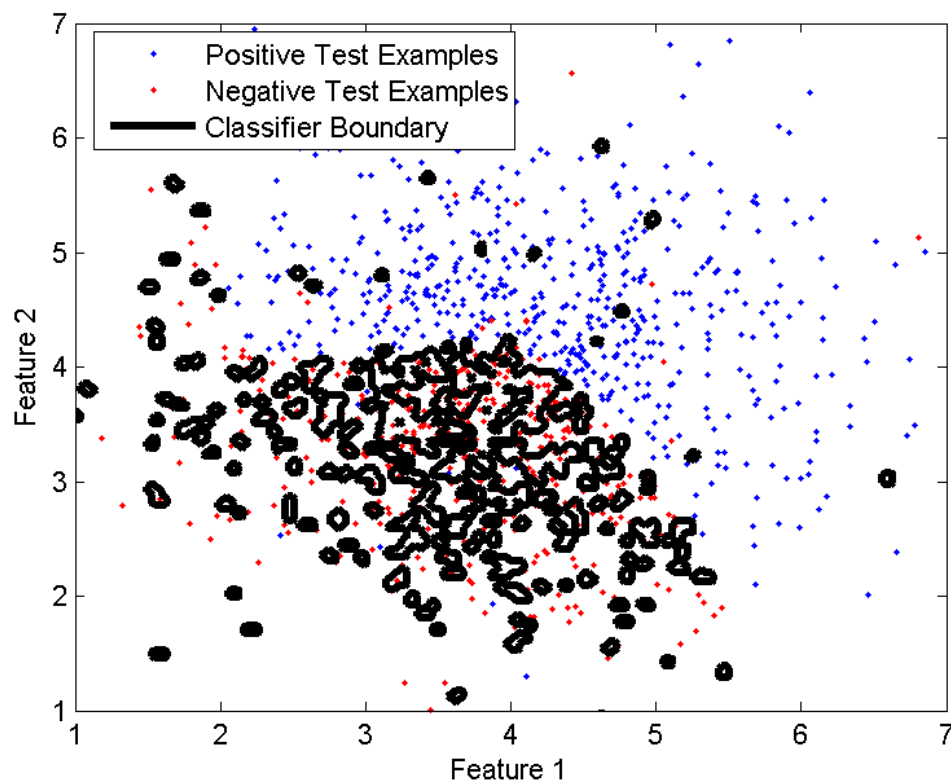
The testing decision boundary is shown below.



- For  $\sigma = 1000$ , the training decision boundary is shown in the plot below.

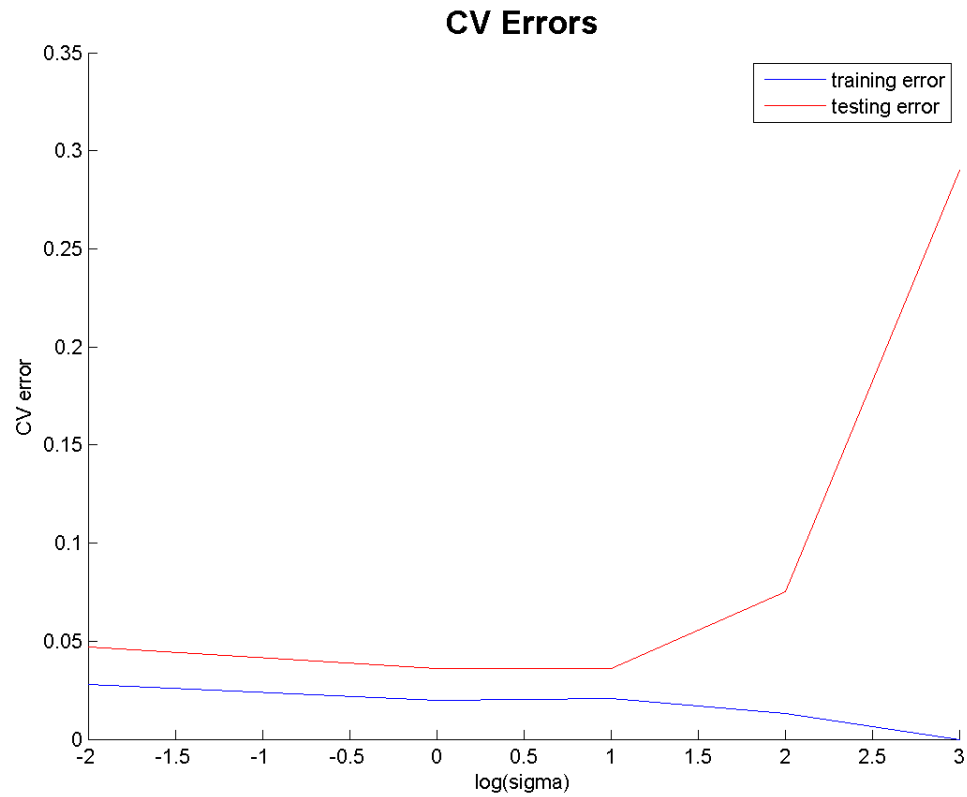


The testing decision boundary is shown below.

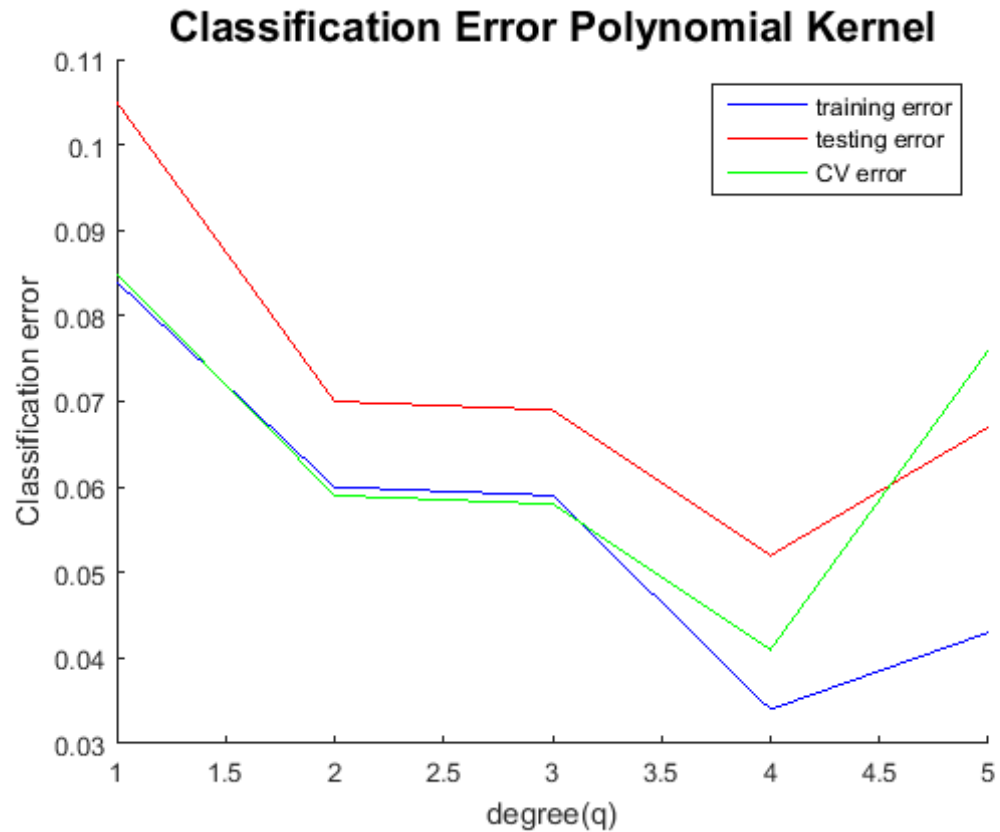


- The training and testing error achieved by each value of  $\sigma$  is shown in the plot below. The value of  $\sigma$  that achieves the lowest test error is  $\sigma = 1$  or  $\sigma = 10$  (equal).

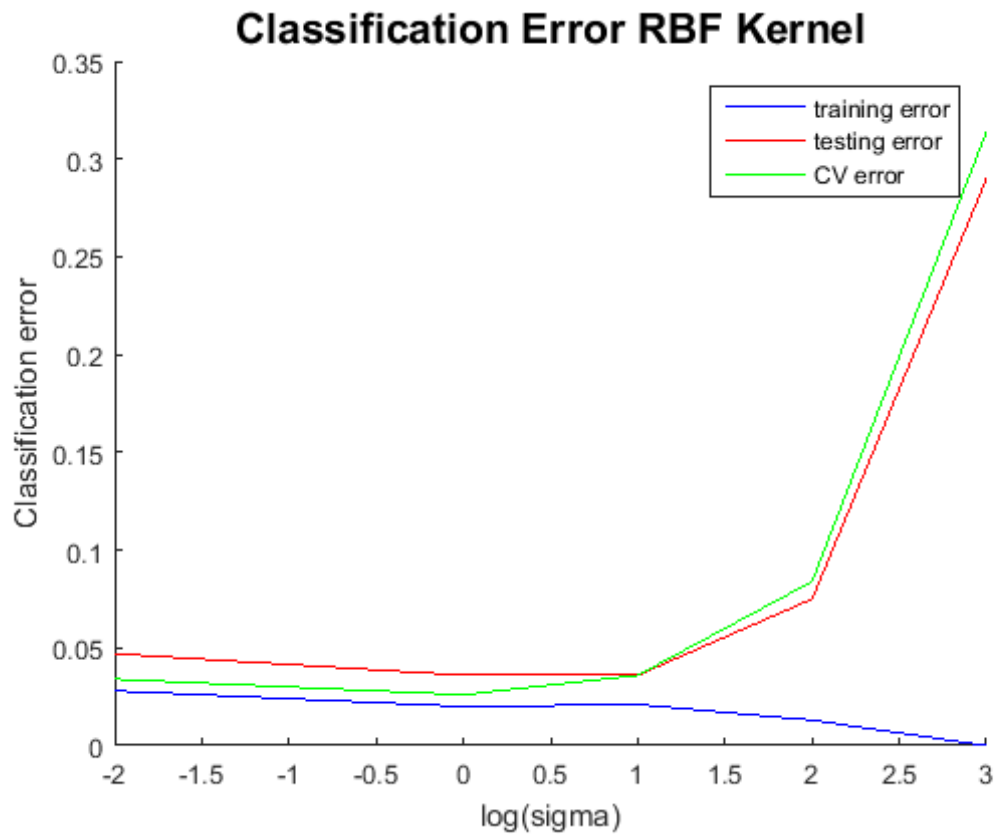




3. • The following plot shows the training, testing, and cross-validation errors for the polynomial kernel. The CV error shows that  $q = 4$  is the best kernel parameter, which corresponds to the lowest test error found in part 1.



- The following plot shows the training, testing, and cross-validation errors for the RBF kernel. The CV error shows that  $\sigma = 1$  is the best kernel parameter, which does correspond to the lowest test error found in part 2. This confirms that cross-validation is a great way to select the kernel parameters.



4. The test accuracy for the polynomial kernel was 0.95181 and 0.96386 for the RBF kernel.