Programming Exercise 6: Support Vector Machines

1.1 Example Dataset 1

Figure1:Example Dataset 1

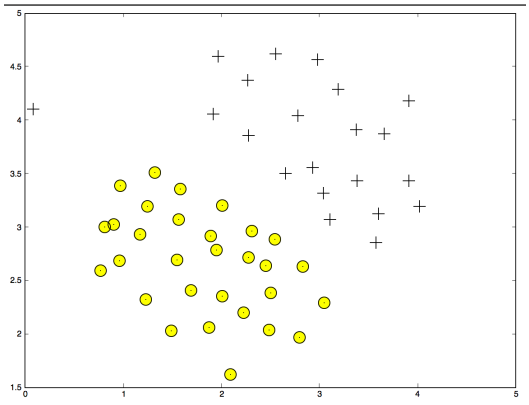


Figure2:SVM Decision Boundary with C = 1(Example Dataset 1)

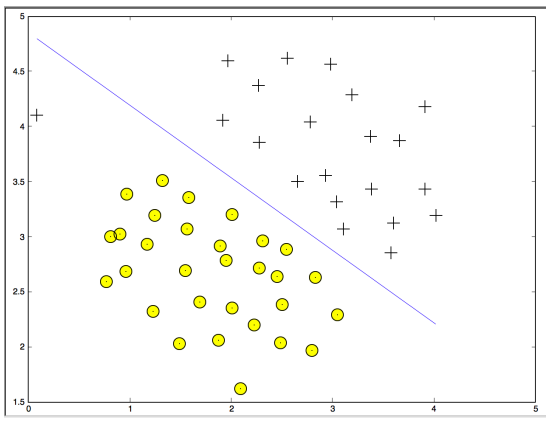
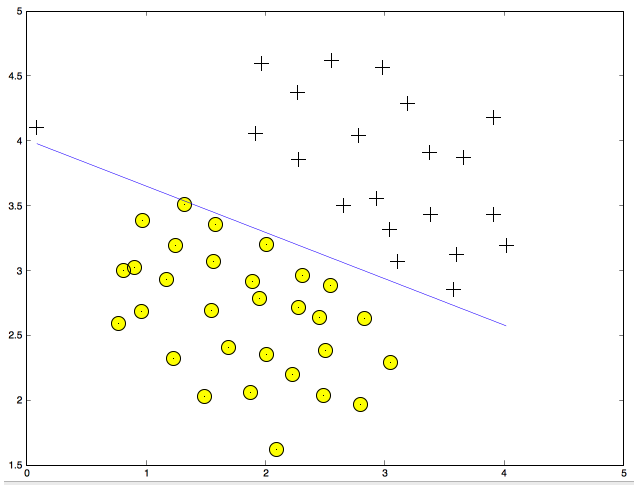


Figure3: SVM Decision Boundary with C =100(Example Dataset 1)



1.2 SVM with Gaussian Kernels

1.2.1 Gaussian Kernel

$$K_{gaussian}(x^{(i)}, x^{(j)}) = \exp\left(-\frac{\|x^{(i)} - x^{(j)}\|^2}{2\sigma^2}\right) = \exp\left(-\frac{\sum_{k=1}^{n}(x_k^{(i)} - x_k^{(j)})^2}{2\sigma^2}\right).$$

```
function sim = gaussianKernel(x1,x2,sigma)
x1=x1(:);
x1=x2(:);
sim=0;
sim=exp(-sum(sum((x1-x2).^2))/2/sigma/sigma);
end
```

1.2.2 Example Dataset 2

Figure 4: Example Dataset 2

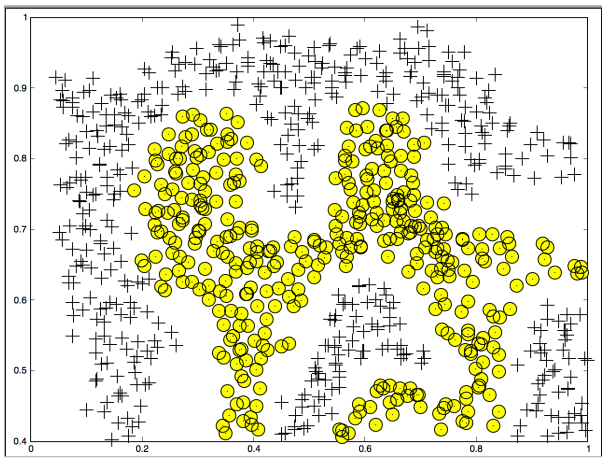
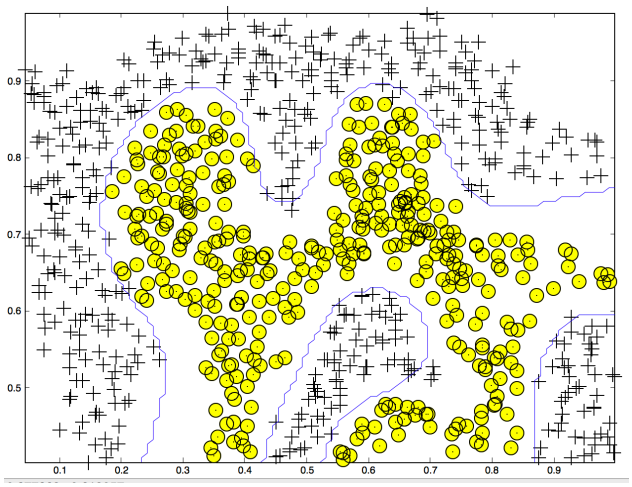


Figure 5:SVM(Gaussian Kernel)Decision Boundary(Example Dataset 2)



1.2.3 Example Dataset 3

Figure 6:Example Dataset 3

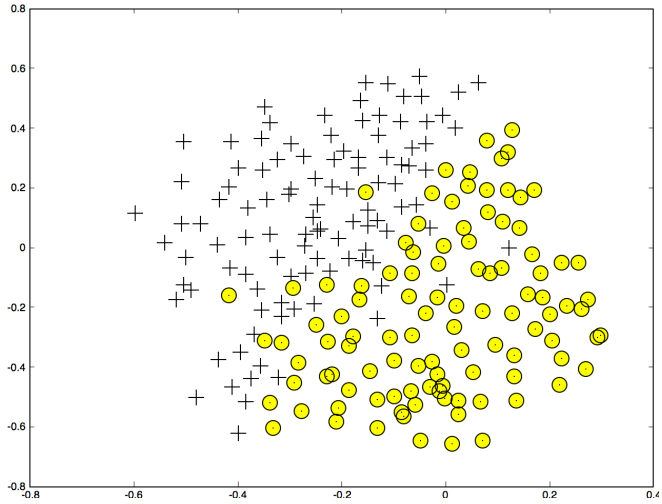
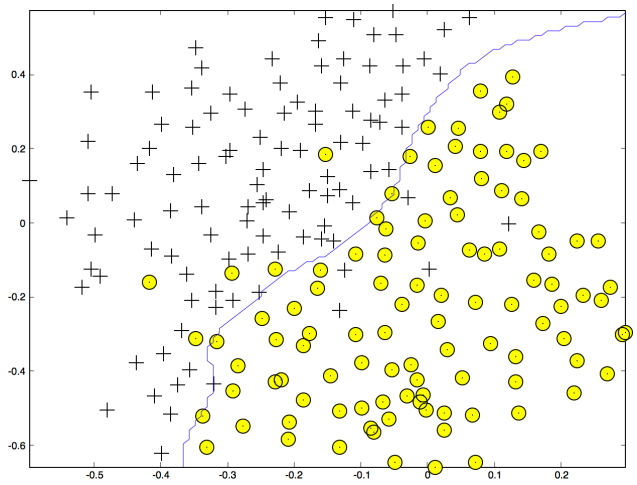


Figure 7:SVM(Gaussian Kernel)Decision Boundary (Example Dataset 3)



```
function [C,sigma] = dataset3Params(X,y,Xval,yval)
C = 1;
sigma = 0.3;

vec = [0.01 0.03 0.1 0.3 1 3 10 30];

for i=1:length(vec),
for j=1:length(vec),
C = vec(i);
sigma = vec(j);
model = svmTrain(X,y,C,@(x1,x2) gaussianKernel(x1,x2,sigma));
predictions = svmPredict(model,Xval);
errors(i,j) = mean(double(predictions~=yval));
end
end

error = min(min(errors));
[i,j] = find(errors == error);
```

```
C = vec(i);
sigma = vec(j);
end
```

2 Spam Classification

2.1 Preprocessing Emails

2.1.1 Vocabulary List

In the code, you are given a string str which is a single word from the processed email. You should look up the word in the vocabulary li vocabulary list. If the word exists, you should add the index of the word into the word indices variable. If the word does not exist, and is word.

```
for i=1:length(vocabList),
if strcmp(str,vocabList)==1,
word_indices = [word_indeics;i];
end
end
```

2.2 Extracing Features from Emails

```
for i=1:size(word_indices,1),
x(word_indices(i))=1;
end
```