ENGG 5202: Assignment #3

Due on Thursday, April 7, 2016

Kai Chen

Problem 1

1.1

$$\begin{aligned}
K_3(x,x') &= K_1(x,x') + K_2(x,x') \\
&= \Phi_1(x)\Phi_1(x') + \Phi_2(x)\Phi_2(x') \\
&= [\Phi_1(x),\Phi_2(x)] \cdot [\Phi_1(x'),\Phi_2(x')] \\
\Phi_3(x) &= [\Phi_1(x),\Phi_2(x)]
\end{aligned}$$

1.2

$$\begin{aligned}
K_3(x,x') &= K_1(x,x')K_2(x,x') \\
&= \Phi_1(x)\Phi_1(x')\Phi_2(x)\Phi_2(x') \\
\Phi_3(x) &= \Phi_1(x)\Phi_2(x)
\end{aligned}$$

1.3

$$\begin{aligned}
K(x,x') &= 1 + x\cdot x' + 4(x\cdot x')^2 \\
&= 1 + x_1x_1' + x_2x_2' + 4(x_1x_1' + x_2x_2')^2 \\
&= 1 + x_1x_1' + x_2x_2' + 4x_1^2x_1'^2 + 4x_2^2x_2'^2 + 8x_1x_1'x_2x_2' \\
\Phi(x) &= [1\ x_1\ x_2\ 2x_1^2\ 2x_2^2\ 2\sqrt{2}x_1x_2]
\end{aligned}$$

Problem 2

2.1

The best kernels and corresponding test errors for different datasets are shown in Table 1. Support vectors are plotted in Figure 1 to Figure 3.

Table 1: Choosing kernel for different datasets

Dataset	Best kernel	Test error
Set1	Linear kernel	4.46%
Set2	Radial basis kernel	1.4%
Set3	Radial basis kernel	0%

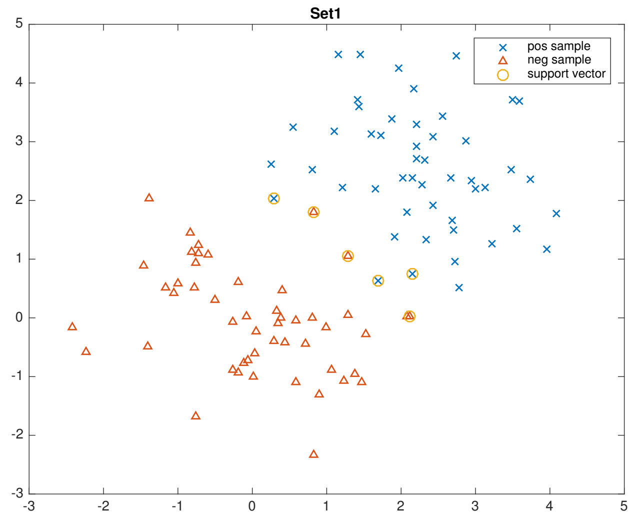

Figure 1: Support vectors of Set1

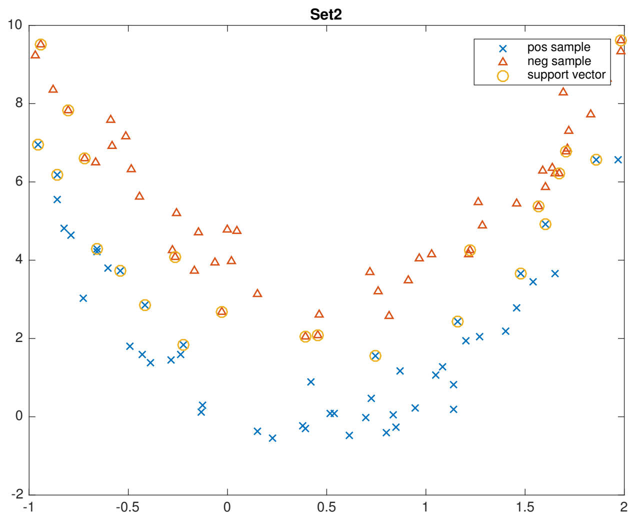

Figure 2: Support vectors of Set2

Table 2: SVM classification error of different kernels

Kernel	Test error
Linear kernel	13.75%
Polynomial kernel	12%
Radial basis kernel	8.5%

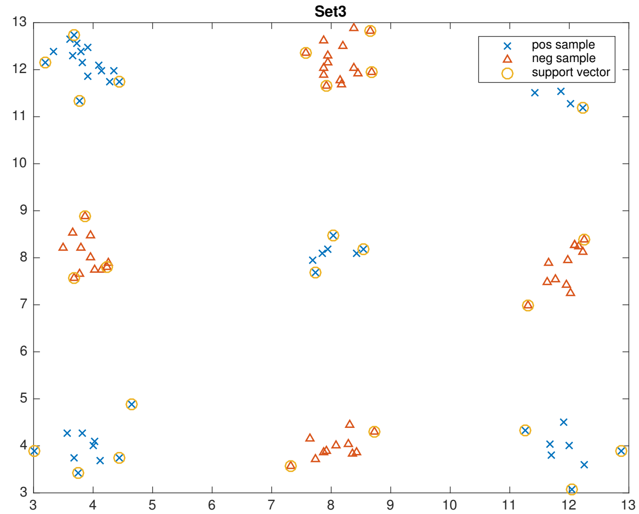

Figure 3: Support vectors of Set3

2.2

Test errors are shown in Table 2.

Problem 3

3.1

The dicision boundary is shown in Figure 4.

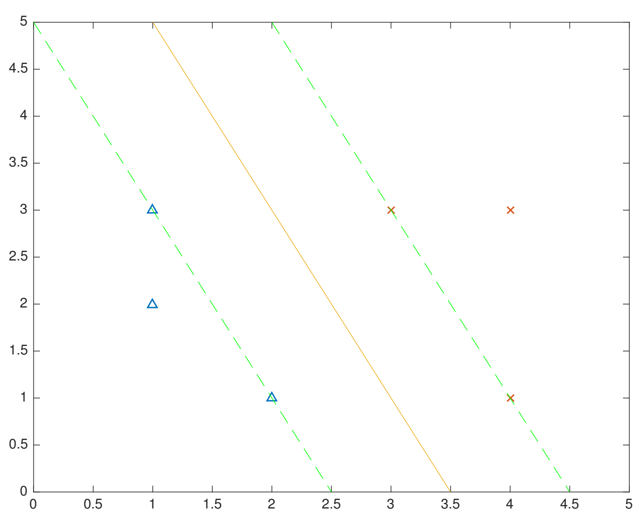

Figure 4: SVM classifier

3.2

Support vectors are these 4 points:
(1, 3)
(2, 1)
(3, 3)
(4, 1)

3.3

Yes, for example, if a negative sample (3, 1) is added, then the number of support vectors will be decreased to 3.

3.4

The leave-one-out cross-validation error is $\frac{1}{3}$ because there are 2 misclassified samples.

Problem 4

4.1

The disicion boundary of the first decision stump chosen by Adaboost is drawn in Figure 5.

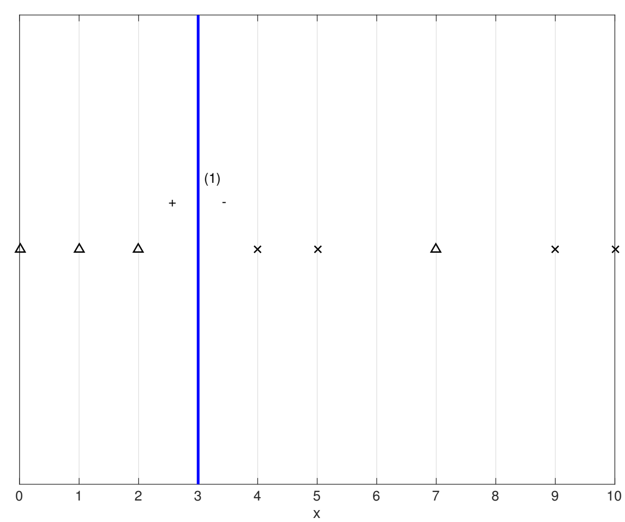

Figure 5: Decision boundary of the first decision stump

4.2

Calculate the vote of the first classifier

$$\begin{aligned}\epsilon_1 &= 0.5 - \frac{1}{2}(\sum_{i=1}^{n} \tilde{W}_i^{(0)} y_i h(x_i; \hat{\theta}_1)) \\ &= \frac{1}{8} \\ \alpha_1 &= 0.5 \ln \frac{1-\epsilon_1}{\epsilon_1} \\ &= 0.5 \ln 7\end{aligned}$$

The sixth sample is misclassified, so

$$\begin{aligned}W_6^{(1)} &= W_6^{(0)} \cdot \exp\{-y_6 \alpha_1 h(x_6; \hat{\theta}_1)\} = \frac{7}{8}\sqrt{e} \\ W_i^{(1)} &= W_i^{(0)} = \frac{1}{8} \quad \text{for all } i \neq 6\end{aligned}$$

After normalization, we can get the new weight for all training samples

$$\begin{aligned}\tilde{W}_6^{(1)} &\approx 0.6225 \\ \tilde{W}_i^{(1)} &\approx 0.0539 \quad \text{for all } i \neq 6\end{aligned}$$

Through line search, we can get the decision boundary of the second decision stump, as is shown in Figure 6.

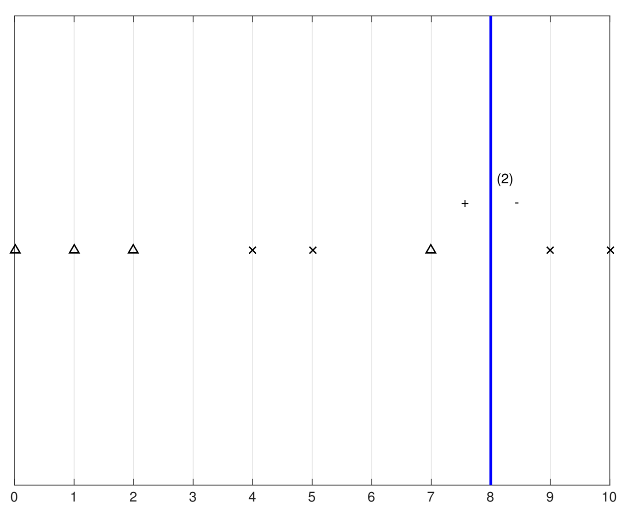

Figure 6: Decision boundary of the second decision stump

4.3

The decision boundary of the first decision stump chosen by Adaboost is drawn in Figure 7.

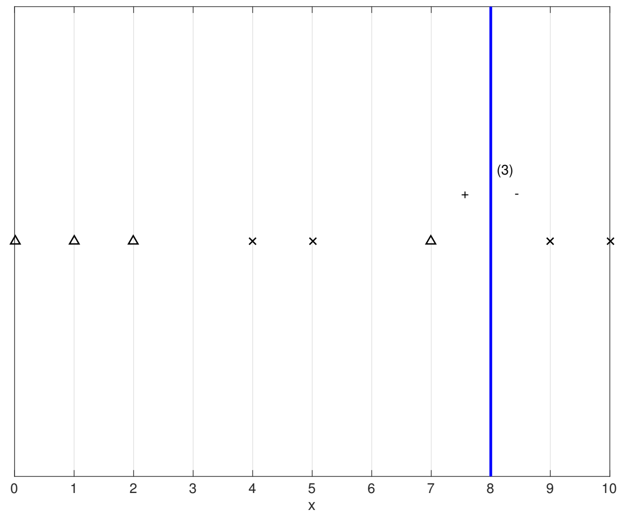

Figure 7: Decision boundary of the first decision stump

4.4

The two finally trained classifiers will behave differently on the test data in terms of classification accuracy, because different initial weight means different data distribution.

4.5

Set the initial weights of positive samples twice of negative samples. This is equivalent to duplicate all the positive samples, which will lead to twice cost of misclassifying a positive sample as a negative sample.