ENGG 5202: Homework #3

Due on Thursday, April 7, 2016, 5:30pm

Xiaogang Wang

Problem 1

[**25 points**]

1. A kernel function $K(\mathbf{x}, \mathbf{x}')$ corresponds to a feature mapping $\mathbf{x} \to \phi(\mathbf{x})$, under which $K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x}) \cdot \phi(\mathbf{x}')$. Let $\phi^{(1)}(\mathbf{x})$ and $\phi^{(2)}(\mathbf{x})$ be the feature vectors corresponding to kernels $K_1(\mathbf{x}, \mathbf{x}')$ and $K_2(\mathbf{x}, \mathbf{x}')$ respectively. According to the kernel construction rules, $K_3(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$ is also a kernel. $K_3(\mathbf{x}, \mathbf{x}')$ corresponds to a feature vector $\phi^{(3)}(\mathbf{x})$. Show that $\phi^{(3)}(\mathbf{x})$ can be expressed in terms of $\phi^{(1)}(\mathbf{x})$ and $\phi^{(2)}(\mathbf{x})$.

2. According to the kernel construction rules, $K_4(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}')K_2(\mathbf{x}, \mathbf{x}')$ is also a kernel. It corresponds to a feature vector $\phi^{(4)}(\mathbf{x})$. Show that $\phi^{(4)}(\mathbf{x})$ can be also derived from $\phi^{(1)}(\mathbf{x})$ and $\phi^{(2)}(\mathbf{x})$.

3. Let $\mathbf{x} = (x_1, x_2)$ be a 2D feature vector. A kernel $K(\mathbf{x}, \mathbf{x}')$ is defined as

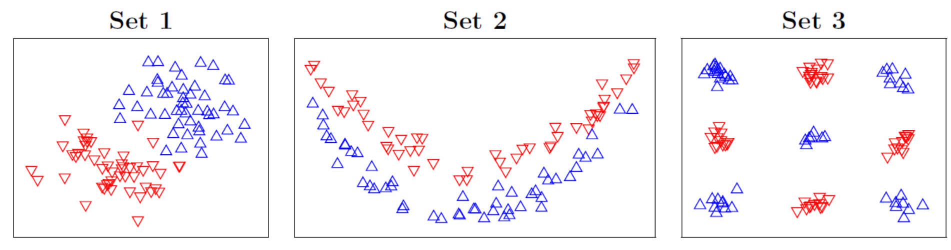
$$K(\mathbf{x}, \mathbf{x}') = 1 + \mathbf{x} \cdot \mathbf{x}' + 4(\mathbf{x} \cdot \mathbf{x}')^2.$$

   What is the feature mapping $\phi(\mathbf{x})$ corresponding to this kernel?

Problem 2

[**30 points**]

LibSVM(http://www.csie.ntu.edu.tw/~cjlin/libsvm/) is one of the widely used libraries for SVM. It has a matlab interface. In this problem, you need to use LibSVM to do some experiments on the data we provided. There are four training/testing sets stored in the **svm.mat** file. The first three data sets are 2D data and they are plotted in the following figure. The fourth is a 64-dimensional digit data set.



- For the first three data sets, consider three kernels: a linear kernel, a second order polynomial kernel and a radial basis kernel ($\sigma = 1$). For each data set, choose the best kernel for a SVM classifier to be trained on the data set. With the chosen kernel for each data set, train the SVM classifier and report the test error. Plot the support vectors. C is fixed as 1000 in all the experiments.
- For the fourth data set, train and test a SVM with a linear kernel, a second order polynomial kernel, and a radial basis kernel ($\sigma = 1.5$). Report the test errors.

Problem 3

[**20 points**]

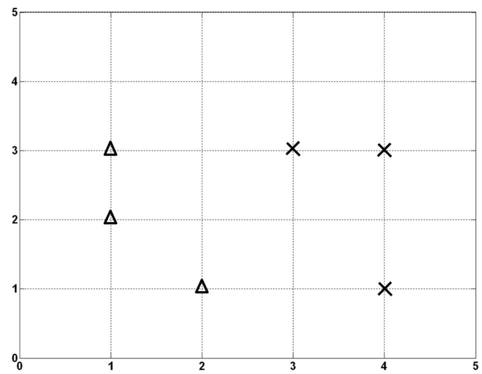


Figure 1: Six two-dimensional data points from two classes, which are indicated by "$\triangle$" and "$\times$".

Consider the six two-dimensional training points plotted in Figure 1. They are labeled into two classes. "$\triangle$" and "$\times$" denote positive and negative samples. Train a linear SVM without slack penalties from these samples.

1. Draw the decision boundary found by SVM and calculate its margin.
2. What are the support vectors?
3. Is it possible to **decrease** the number of support vectors by adding an extra training example? If yes, show how to do it? If no, explain the reason.
4. What is the leave-one-out cross-validation error of the linear SVM without slack penalties trained on this training set?

Problem 4

[**25 points**]

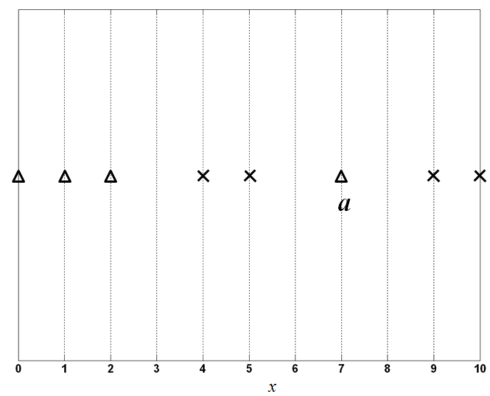


Figure 2: Eight one-dimensional data points from two classes, which are indicated by "$\triangle$" and "$\times$".

Consider the eight one-dimensional training points plotted in Figure 2. They are labeled into two classes. "$\triangle$" and "$\times$" denote positive and negative samples. Adaboost with decision stumps is applied to this training set.

1. In initialization, the weight for every data point is set as $\tilde{W}_i^{(0)} = 1/8, i = 1, \ldots, 8$. Draw the decision boundary of the first decision stump chosen by Adaboot. Indicate the $+/-$ side of the decision boundary. Label this decision boundary as (1).

2. Draw the decision boundary of the second decision stump chosen by Adaboot. Indicate the $+/-$ side of the decision boundary. Label this decision boundary as (2).

3. Now consider a different initialization, where the weight for data point "a" is set as $3/10$ and the weights for all the other points are $1/10$. With this initialization, draw the decision boundary of the first decision stump chosen by Adaboot. Indicate the $+/-$ side of the decision boundary. Label this decision boundary as (3).

4. The two different initializations in question 1 and question 3 lead to different combined classifiers after the training processes of Adaboost complete. Will the two finally trained classifiers behave similarly or differently on the test data in terms of classification accuracy? In other words, is Adaboost sensitive to the initialization of training weights $\tilde{W}_i^{(0)}$? Explain your answer.

5. In the Adaboost algorithm taught in class, we assume that both types of classification errors have equal cost. However, it may not always be the case in some practical applications. Consider that the cost of misclassifying a positive sample as a negative sample is twice as that of misclassifying a negative sample as a positive sample. Please modify the Adaboost algorithm to incorporate unequal costs.