Problem 2

Data set: 337 data with 12 attributes.

(a)

I did five cross validation on the data set, and repeat the experiment 5 times, then recording the results as following.

Neural network I used:

10 hidden nodes with 'logsig' function, one output neuron with linear function 'purline'.

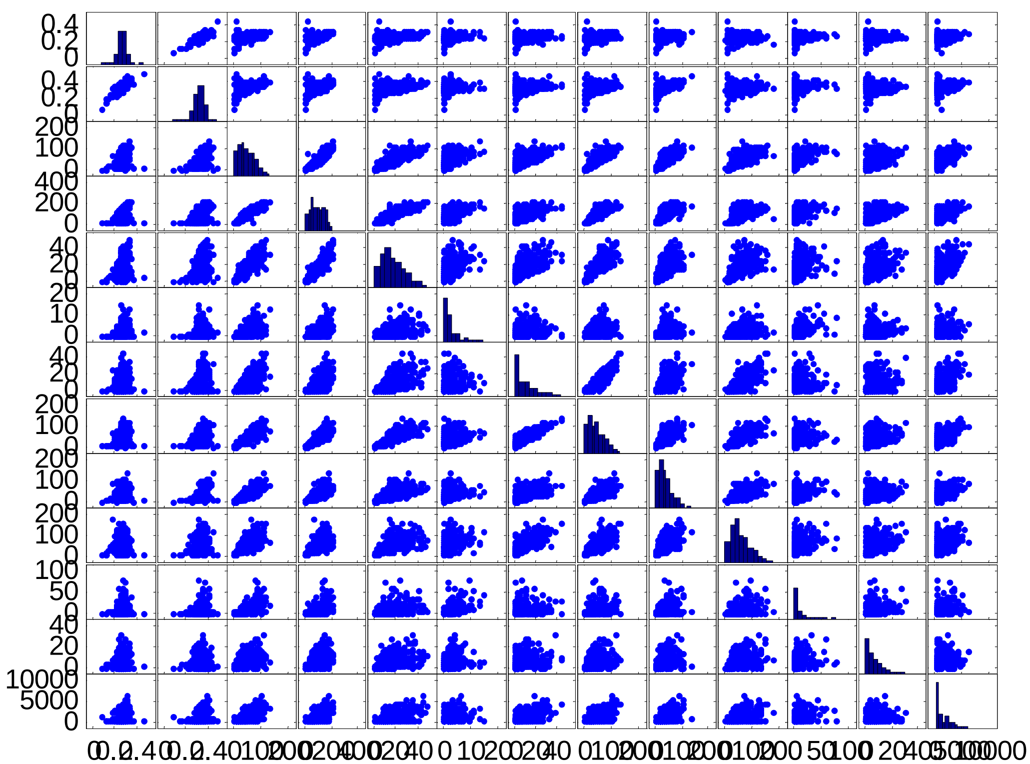
Maximum epochs: 200

R_ square	0.3526	0.4256	0.3825	0.4047	0.3870

Mean R_square: 0.3905

(b)

First, I try to remove the redundant features. The correlation between each pair of features is shown in the following figure:



As we can see in the figure, some features are high correlated each other. First, I remove some of these features, then retrain the model to see whether removing such high correlated features can improve the accuracy.
I removed 3 redundant features. The new accuracy is about 0.3, which means removing redundant features does not help to improve the accuracy.

Next step, I try to remove some outlier.
I apply an outlier detection algorithm (downloaded from: http://www.mathworks.com/matlabcentral/fx_files/11106/1/outlier.m) to the data.
After applying the algorithm, I remove 20 outliers. Than I retrain the model, the R square I get now is also about 0.3.

Conclusion, using all data with all attributes can result in a higher accuracy, since in my experiments neither removing high-correlated features nor removing outliers can improve the accuracy.

Problem 3
Data set: 196 attributes, 2500 data.

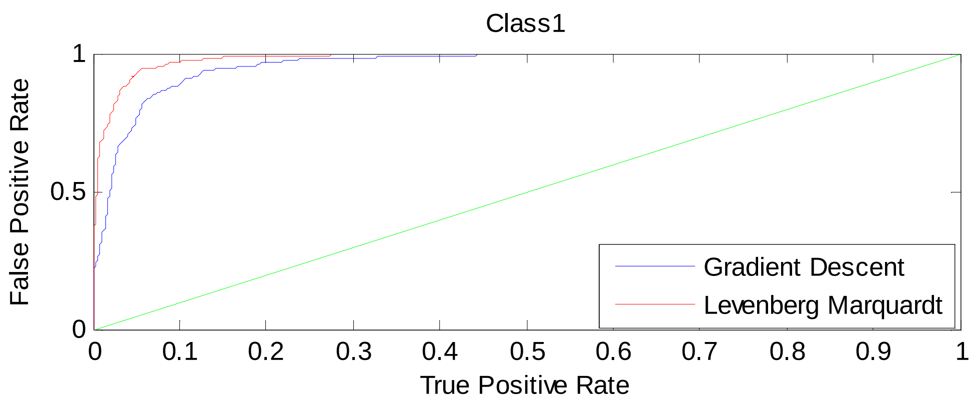
First, I remove the attributes with only zero value, since such attributes will not contribute to build the model. After this preprocessing, the final data set has 174 attributes and 2500 data.

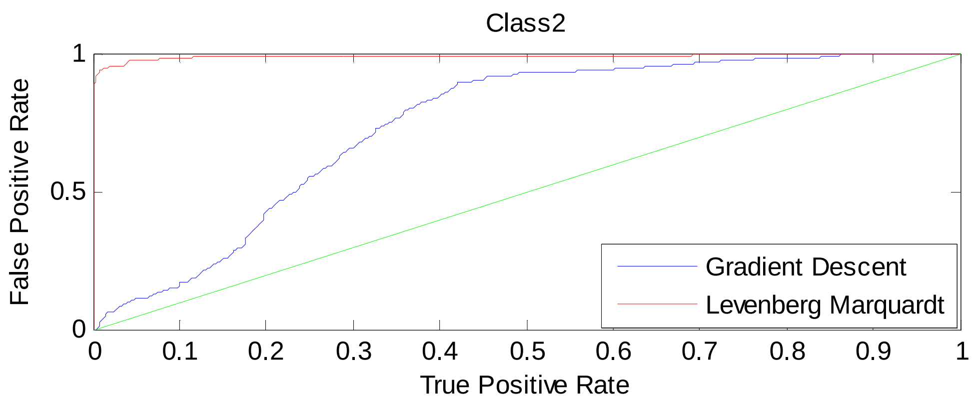
(a) I use two training algorithm here. First one is 'traingd' which is using Gradient Descent method to optimize the weights. The other one is 'trainlm' which is using Levenberg Marquardt algorithm.

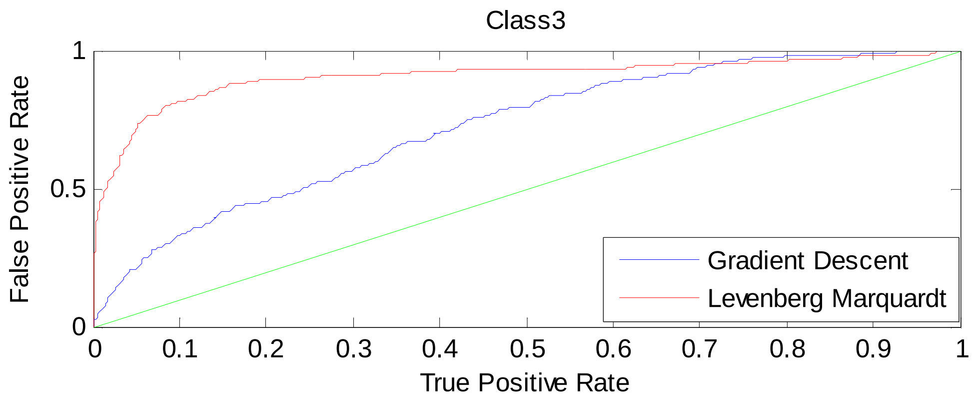
The comparison between two methods in terms of number of epochs, running time and accuracy is listed in the following table.

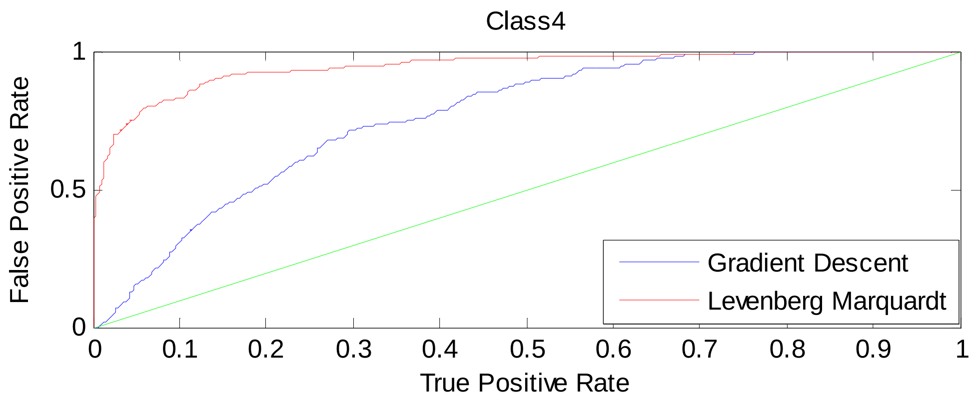
Method	Epochs	Running time	Accuracy
Gradient Descent	700	55.48	0.35
Levenberg Marquardt	15	169.9	0.75

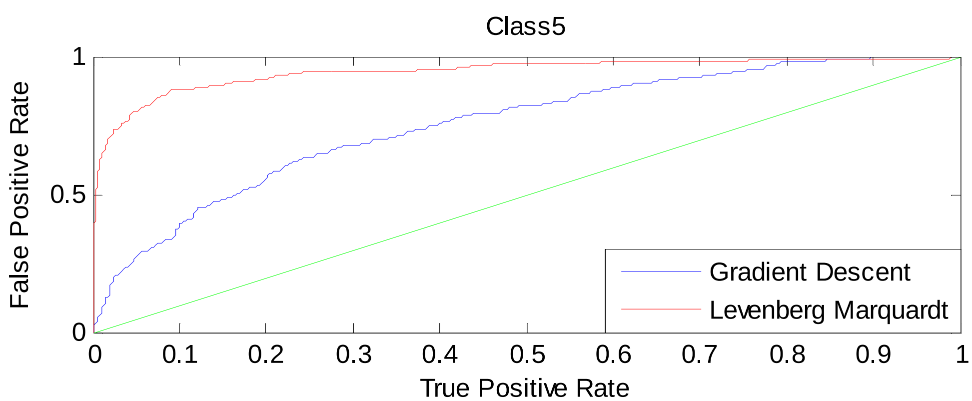
Then I plot the Roc curve of these two methods. I plot 10 figures, each figure corresponds to the Roc curve of one class. Please see the following figures.
From these figures we can also conclude, training method using a second-order optimization methods is much better than gradient descent on this baseball data set.

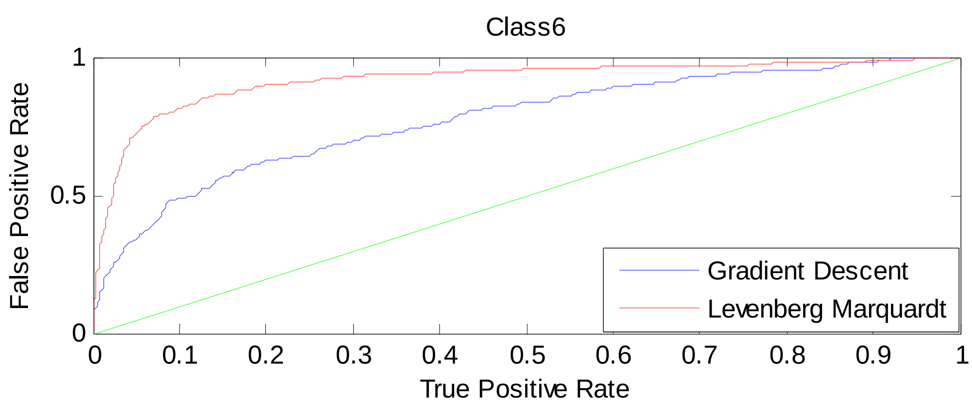


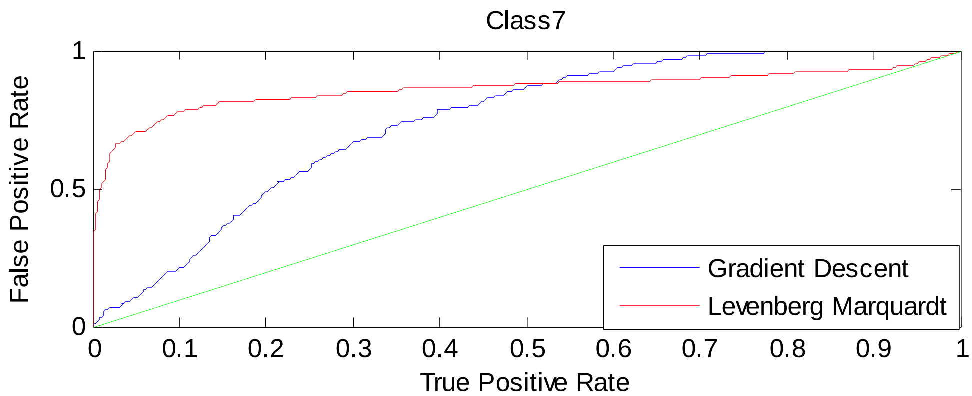


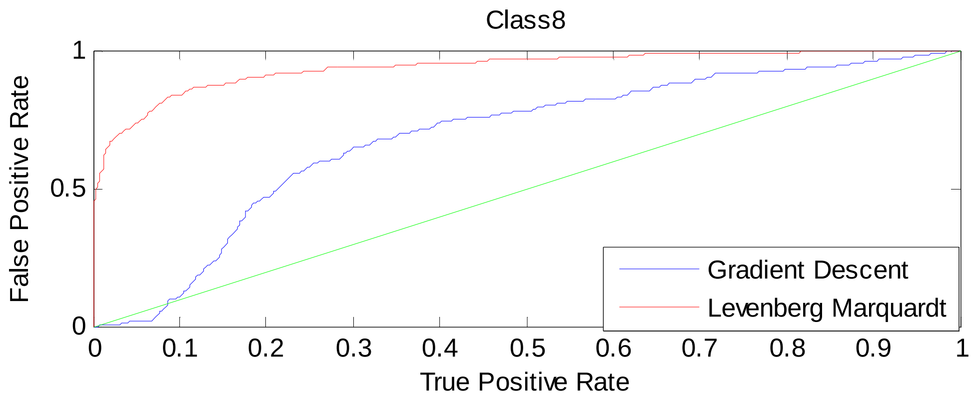

Class3
False Positive Rate
True Positive Rate
Gradient Descent
Levenberg Marquardt

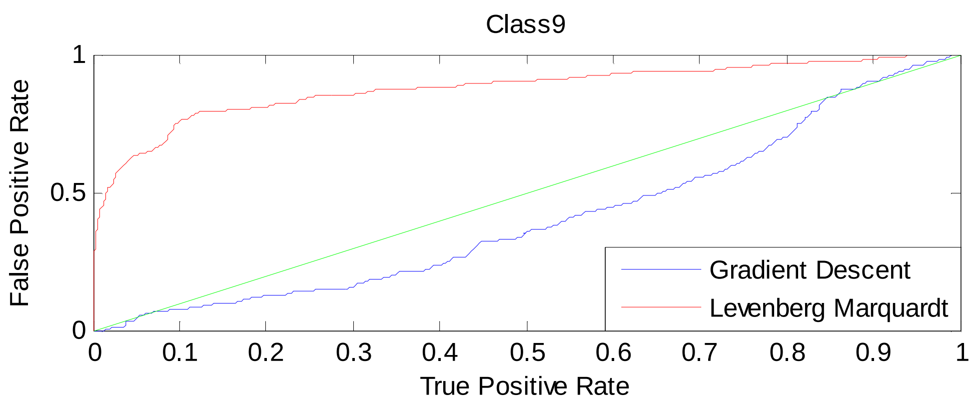

Class4
False Positive Rate
True Positive Rate
Gradient Descent
Levenberg Marquardt

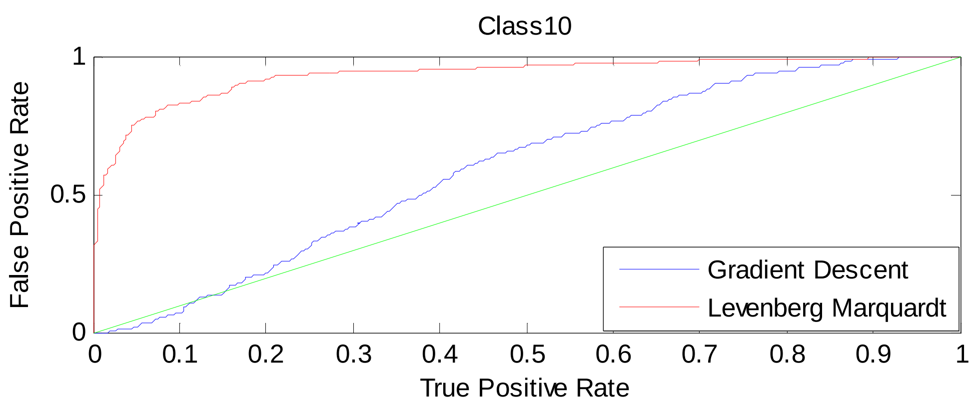

Class5
False Positive Rate
True Positive Rate
Gradient Descent
Levenberg Marquardt


Class6
False Positive Rate
True Positive Rate
Gradient Descent
Levenberg Marquardt


Class7
False Positive Rate
True Positive Rate
0
0.5
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Gradient Descent
Levenberg Marquardt


Class8
False Positive Rate
True Positive Rate
0
0.5
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Gradient Descent
Levenberg Marquardt


Class9
False Positive Rate
True Positive Rate
0
0.5
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Gradient Descent
Levenberg Marquardt


Class10
False Positive Rate
True Positive Rate
0
0.5
1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Gradient Descent
Levenberg Marquardt

(b) and (c)

In these two parts, I tried two things.
First one is building a Radial Basis Function network to build the prediction model, the other one is apply SVM with polynomial kernel. The comparison among these two methods and the methods applied in problem (a) is listed in the following table:

Methods	Running time (s)	Accuracy
Gradient Descent	55.48	0.35
Levenberg Marquardt	169.9	0.75
Radial Basis	22.01	0.83
SVM	96.53	0.86

Conclusion:
1. Second-order optimization is time consuming which is what we can expect and it can get higher accuracy than Gradient Descent methods.

2. SVM and Radial Basis are better choice in such classification problem, especially SVM, it gets highest accuracy.

3. However, here the size of our data is not to big (2500 data), if we have huge data set, SVM is not a good choice as Radial Basis method, since SVM is more time consuming.