Lab 05 – Classification Using Linear Regression

For this lab you will need to format the iris and diabetes training files in a format that is suitable for using the linear regression method by changing the class values into numbers.

Exercise 1 diabetes dataset

1. Make a copy of the file ‘diabetesTrainingSet’. Open the file with a text editor. Replace the class values with numeric values, e.g., replace ‘tested_positive’ with the integer 1 and ‘tested_negative’ with 0. Remember to update the class attribute type to numeric.
2. Fire up Weka, open the diabetes training data set which has integers as class values.
3. Apply the linear regression algorithm to obtain an equation that fits the data.
4. Open the excel file diabetesTestSet, apply the equation to all the instances. With the results determine the class of each instance. What is your evaluation of the linear regression method?
5. How well does it do when compared with other classification algorithms that you have tried before?

Exercise 2 iris dataset

1. Make multiple copies of the file ‘irisTrainingSet’. Use a text editor to prepare the files so that the multi-response linear regression method for classification can be applied.
2. Fire up Weka and apply the linear regression algorithm to the three files to obtain 3 equations.
3. Open the excel file irisTestSet, apply the equations to all the instances. With the results determine the class of each instance.

What is your evaluation of the multi-response linear regression method?

4. How well does it do when compared with other classification algorithms that you have tried before?