# Machine Intelligence

# Artificial Intelligence Techniques

# Tutorial

# Iris Data and Classifiers

# Semester 1 2015

### Part A) The Iris Data set

Today we will be using iris.dat which is available from the subject web site.

A description of the Iris data is:

|  |
| --- |
| I have sourced this dat set and relevant information from: <http://www.ics.uci.edu/~mlearn/MLRepository.html>  This data set was first identified and used by: R.A. Fisher in his paper: “The use of multiple measurements in taxonomic problems (1936)”  This is one of the best known database to be found in the pattern recognition literature.  Fisher's (1936) paper is a classic in the field and is referenced frequently to this day.  It contains 150 instances, each has four measurements which are:  col 1. sepal length in cm  col 2. sepal width in cm  col 3. petal length in cm  col 4. petal width in cm  There is one class value between 1 and 3  1 Iris Setosa  2 Iris Versicolour  3 Iris Virginica |

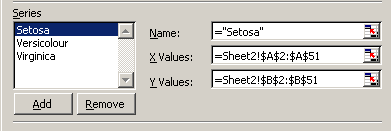
Fragment of iris.dat:

|  |
| --- |
| 0.50 0.33 0.14 0.02 1  0.64 0.28 0.56 0.22 3  0.65 0.28 0.46 0.15 2  0.67 0.31 0.56 0.24 3  0.63 0.28 0.51 0.15 3  0.46 0.34 0.14 0.03 1  0.69 0.31 0.51 0.23 3  0.62 0.22 0.45 0.15 2  0.59 0.32 0.48 0.18 2 |

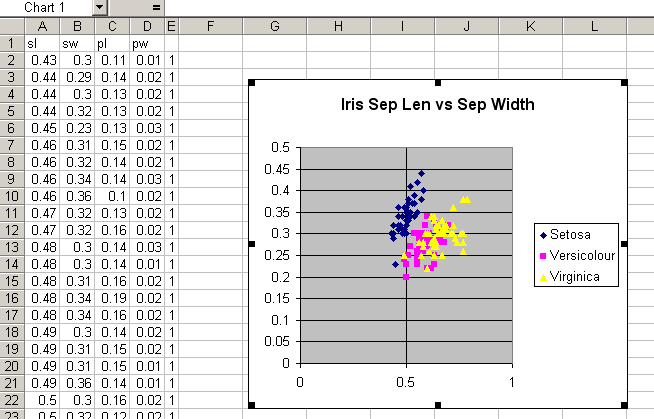
As you may expect this data has been scaled so the first 4 columns are between 0 and 1.

You need to run up an excel spreadsheet and graph the first two columns of data with different colors for the different values in the final column.

In order to do this you need to sort the data by column 5.



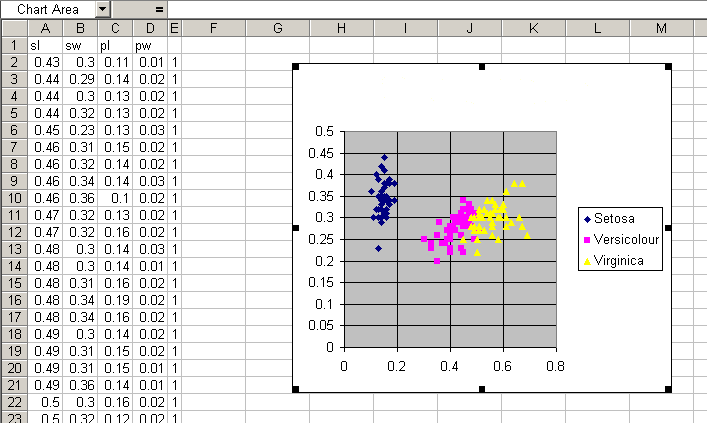
You graph (which needs 3 data series) should look something like this:



Note how the classifications of Versicolour and Virginica overlap it would be difficult to draw a line between them.

Now graph cols 1 and 3 (sepal length vs petal length)

it should look something like this:



Note how the classifications of Versicolour and Virginica overlap less it would be possible to draw a line between them.

Based on the above graphs one could construct a set of rules to identify whether set of readings come from a Setosa, Versicolour or Virginica Iris plant.

eg if Col3 (petal length) is less than 0.2 then the plant is a Setosa Iris.

Next week we will train a neural network to process this data and do exactly that it will be able to predict for four columns of numbers whether the plant is a Setosa, Versicolour or Virginica Iris plant. This week however try constructing a few rules in excel to do the job. Measure how accurate you are based on how many it gets right.