**WEKA Machine Learning: Naive Bayes**

**Aim**: Predict whether tomorrow’s weather will be good to play football using a sample of 14 weather samples based on outlook, temperature, humidity, wind and when the game did play

**Data**: weather.numeric

**Algorithm**: Naive Bayes

The algorithm is the n-gram process and calculates probabilities.

・ Load file in the **Preprocess** tab

◦ Examine the data with the **Edit** button in **Preprocess** – you will see a table of the data

◦ Visualise the unprocessed data with **Visualize All** on the **Process** tab – you will see graphs of each attribute. Hover over the graph bars to see what they represent.

▪ Look at the attributes, type, properties, how may there are in that class

・ Select a Bayes algorithm in the Classify tab

◦ Bayes > BayesNet – this implements Naive Bayes

▪ You can click in the command line for the classifier to change parameters

▪ NB there are some further options on More Options – which we do not need today

・ Test our Classifier in the Classify tab

◦ There are 4 **Test Options** and we will use **Cross-validation** **Folds 10**

▪ **Use training set** means that you use the training set (the file you loaded in Preprocess) for testing.

▪ **Supplied test set** means that you can specify a file with the test data. To do this you select the option and click on **Set…**Now you get a small window called **Test Instances** that allows you to load a test file and then shows you the name of the relation and the number of attributes and instances. Click on **Open file…** and load the test file (e.g. **weather-test.arff**). You get the following:

▪ **Cross-validation** means that the classification results will be evaluated by cross-validation. In this mode you can also change the number of folds.

▪ **Percentage split** means that classification results will be evaluated on a test set that is a part of the original data. The default split (shown in the text area next to the option) is 66%, which means that 66% of the data go for training and 34% for testing.

◦ Click the **Start** button and a window will open with the results.

▪ Under Start you will see Bayes, Bayesnet.

・ To view the network right-click on Bayes, Bayesnet and you will see a graph of the data.

・ To see the data that results, right-click on Bayes, Bayesnet and then right-click on Bayes, Bayesnet visualise Classifier Errors and the Save button. Save the file and then load it into Weka with Preprocess and view it with Edit.

◦ You will see the values that have been given to each of the 14 days with predicted play being the result of the guesses made by the naive Bayesian algorithm.

Explore all the options to Visualise (right click on **Start**)