CFCS Tutorial Two

Miles Osborne

January 21, 2008

This tutorial deals with simple vector-based machine learning. Suppose you are given the following sets of reading about *dogs*:

1	0.75	0.25
2	0.75	0.5
3	0.5	0.5

Also, you have the following readings about *cats*:

1	0.5	0.6
2	0.25	0.5
3	0.25	0.75

1. Using a *Knn* classifier, work-out whether the following examples are *cats* or *dogs*:

1	0.6	0.6
2	0.75	0.25
3	0.25	0.25

2. Now, using some paper, create a two-d graph and mark the various points on it. Do your previous results agree with a visual inspection?

3. What happens as you vary k?

4. If you saw 10 more of the type two *cat* reading, what would happen to your results?

5. Suppose a *cat* reading is the same as a *dog* reading. What would happen?

6. Our set of *cat* and *dog* readings very usefully told us which kind of animal went with which kind of reading. This is called *supervised machine learning*. *Unsupervised machine learning* deals with examples that have no explicit label –we do not know whether a set of readings came from a *dog* or a *cat*. How could you assign labels to such readings? You should think about what would happen if you added these newly labelled examples to your initial set of labelled examples.