

341 Introduction to Bioinformatics:

Biological Networks

Tutorial, week 9 – March 18, 2010

1. Run the nearest neighbour clustering on the following points in Euclidean 2-dimensional space using Euclidean distance and display all intermediate steps and the final clustering:
A(3,4), B(3,6), C(3,8) D(4,5), E(4,7), F(5,1), G(5,5), H(7,3), I(7,4), J(8,5)
2. Use single-link, complete-link, average-link hierarchical clustering on the same set of points as above. (Can be done at your own time at home, for practice.)
3. *k-Nearest neighbours*. Given the training data below (*Buy Computer* data), predict the class of the following new example using k-Nearest Neighbour for k=5: age<=30, income=medium, student=yes, creditrating=fair. For distance measure between neighbours use a simple match of attribute values: distance(A,B)=

$$\sum_{i=1}^4 w_i * \partial(a_i, b_i) / 4$$

where $\partial(a_i, b_i)$ is 1 if a_i equals b_i and 0 otherwise. a_i and b_i are either *age*, *income*, *student* or *credit_rating*. Weights are all 1 except for income it is 2.

<i>RID</i>	<i>age</i>	<i>income</i>	<i>student</i>	<i>credit_rating</i>	<i>Class: buys_computer</i>
1	<=30	high	no	fair	no
2	<=30	high	no	excellent	no
3	31 ... 40	high	no	fair	yes
4	>40	medium	no	fair	yes
5	>40	low	yes	fair	yes
6	>40	low	yes	excellent	no
7	31 ... 40	low	yes	excellent	yes
8	<=30	medium	no	fair	no
9	<=30	low	yes	fair	yes
10	>40	medium	yes	fair	yes
11	<=30	medium	yes	excellent	yes
12	31 ... 40	medium	no	excellent	yes
13	31 ... 40	high	yes	fair	yes
14	>40	medium	no	excellent	no

4. **ROC Curves.** Let the following three points have the following sensitivity and specificity:

point	sensitivity	specificity
1	0.56	0.99
2	0.78	0.81
3	0.91	0.42

Draw the ROC curve for these points. Give a definition of a ROC curve and explain what it demonstrates.