COMS20011 – Data-Driven Computer Science

Problem Sheet MM04

1 – What does dimensionality reduction reduce?

A. Stochastics
B. Collinearity
C. Performance
D. Entropy
E. All of the above

2 – In PCA, what is the relationship between the number of principal components retained and the amount of variance explained by the retained components?

A. More retained components explain less variance.
B. More retained components explain more variance.
C. The number of retained components does not affect the explained variance.
D. The relationship depends on the type of dataset.
E. All of the above.

3 – Imagine you have received a huge shipment of three varieties of fruits consisting of *Oranges*, *Satsumas*, and *Red Pears*. The fruit is unfortunately mixed up, but you have access to a vision system you can program to distinguish and separate the fruit as they pass in front of a camera on a converyor belt one at a time. The camera is positioned to give a top-view of the fruit.

(a) State no fewer than three features you would use in your design to distinguish between the different types. Very briefly explain why your features will pick the correct type each time considering that some measurements maybe somewhat affected by noisy data from the image acquisition process.

(b) Consider you had actually been asked to consider using up to 20 features for this task. Discuss what would you do to find out which features are significant (or which ones are redundant)?

4 – Given a feature set X_i, i = 1, ..., 20, we wish to find a subset *Y* of 9 features, that gives the best *P(correct classification)*. Determine how many times the *P(correct classification)* function will have to be evaluated to find the best subset *Y*.

5 – The eigenvalues of a 6D dataset are: **[17, 11, 8, 2, 0.65, 0.35]**. What *variance* in the dataset do the first 3 eigenvalues represent?

(a) 93.2%
(b) 91.3%
(c) 96.3%
(d) 92.3%

6 – Calculate the result of the convolution $A*B$ in each of the examples below by hand.

(i) $A = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix}$ $\quad B = \begin{pmatrix} 2 & 2 & 3 & 3 & 2 \end{pmatrix}$

(ii) $A = \begin{pmatrix} 1 & 1 & 3 & 1 & 1 \end{pmatrix}$ $\quad B = \begin{pmatrix} 3 & 3 & -2 & -1 & 0 & 1 & 2 & 3 & 3 \end{pmatrix}$

(iii) $A = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$ $\quad B = \begin{pmatrix} 0 & 5 & 5 & 5 & 0 \\ 0 & 5 & 10 & 5 & 0 \\ 0 & 10 & 10 & 10 & 0 \\ 0 & 5 & 10 & 5 & 0 \\ 0 & 5 & 5 & 5 & 0 \end{pmatrix}$

Now verify your result using Matlab (using *conv* and use *help conv* to determine what convention Matlab uses when convolving at the border points) or using Python.

7 – A 3x3 spatial filter with all elements set to **−1,** except the central element set to **16,** has a...

(a) normalisation factor of *1/8*
(b) normalisation factor of *1/16*
(c) normalisation factor of *1/24*
(d) normalisation factor of *1/12*

8 – Apply filter M to matrix K at location (4,3), once using 4-connectivity and then 8-connectivity.

$$M = \begin{pmatrix} 1 & 1 & -1 \\ 4 & 8 & -4 \\ 1 & 1 & -1 \end{pmatrix} \qquad K = \begin{pmatrix} 10 & 8 & 0 & 0 & 1 \\ 8 & 9 & 2 & 1 & 0 \\ 8 & 8 & 1 & 1 & 2 \\ 6 & 7 & 1 & 1 & 3 \\ 4 & 8 & 2 & 0 & 3 \end{pmatrix}$$