Dimension Reduction, AMSI 2021 Winter School

Tutorial 2

Anastasios Panagiotelis

July 15, 2021

Images Data - Cleaning

1. Using the images data construct a matrix where each row corresponds to an image and each column to a color channel (R, G or B) of an individual pixel.
2. Find all variables which have no variation across the images and remove them.
3. Convert the matrix to a data frame and add a column to your dataframe with the name of each file.

Images Data - Dimension Reduction

1. Carry out PCA and ISOMAP using the `dimRed` package (note some commands may take about 20-30 second to run - be patient!). Use the function defaults for output dimension and tuning parameters.
2. For PCA and isomap, plot a scatterplot.

Local continuity meta criterion

1. Compute the local continuity meta criterion (LCMC) for PCA and Isomap. Compare the LCMC across all both methods when the number of nearest neighbours used to compute the LCMC is 50. (Note this can be a bit slow)
2. Carry out PCA using 3 PCs and recompute the LCMC.
3. Carry out isomap using 20 nearest neighbours (and two output dimensions) and recompute the LCMC.

coRanking;

4. Compare the LCMC from your answers in 1-3. What do you conclude?

Interpreting the output dimensions

1. Using Isomap computed using 20 nearest neighbours, find the images corresponding to (i) the largest value of the first output dimension (ii) the smallest value of the second output dimension (iii) the largest value of the second output dimension. Judging, from this what may each output dimension represent?

Bonus Question

1. Let $\mathbf{x}$ be a p-vector and $\tilde{\mathbf{X}}$ be a $p \times k$ matrix whose columns are the k nearest neighbours of $\mathbf{x}$

Prove that the value of $\mathbf{w}$ that minimises

$$||\mathbf{x} - \mathbf{w}'\tilde{\mathbf{X}}||_2^2$$

subject to $\sum w_j = 1$ is given by

$$\mathbf{w} = \frac{\sum_k \mathbf{C}^{-1}_{jk}}{\sum_k \sum_l \mathbf{C}^{-1}_{jk}}$$

where $c_{jk} = ||\mathbf{x} - \mathbf{x}_j||_2^2$ and $\mathbf{x}_j$ is a column of $\tilde{\mathbf{X}}$.

2. Let $\mathbf{y} = (y_1, \ldots, y_n)'$ be an n-vector. Show that

$$\sum_i (y_i - \sum_j w_{ij} y_j)^2$$

is minimised subject to $\sum y_i^2 = 1$ by the eigenvector corresponding to the second smallest eigenvalue of $(\mathbf{I} - \mathbf{W})'(\mathbf{I} - \mathbf{W})$.