Learning from Data, Tutorial Sheet for Week 3

School of Informatics, University of Edinburgh

Instructor: Amos Storkey

1. Consider the state of all boolean (true false) two category hypotheses (eg (F ?) and (T T)). Construct a diagram writing out all the hypotheses, and draw arrows from each hypothesis to all hypotheses which are the next most specific according to the general-specific partial ordering.
1b. Suppose we see a negative example (T T). Circle the hypotheses which are elements of the new version space. 1c. Now suppose we receive a positive example (F T). Put boxes around the remaining elements of the version space. Does anything different happen if the data arrived in the reverse order.

2. Consider the following 3-dimensional datapoints:

$(1.3, 1.6, 2.8)(4.3, -1.4, 5.8)(-0.6, 3.7, 0.7)(-0.4, 3.2, 5.8)(3.3, -0.4, 4.3)(-0.4, 3.1, 0.9)$

Perform Principal Components Analysis by:

- Calculating the sample mean, $\mathbf{c}$ of the data.
- Calculating the sample covariance matrix $S = \frac{1}{5}\sum_{\mu=1,\ldots,6} \mathbf{x}^\mu (\mathbf{x}^\mu)^T - \mathbf{c}\mathbf{c}^T$ of the data.
- Finding the eigenvalues and eigenvectors $\mathbf{e}^i$ of the covariance matrix. (This is easy to do in MATLAB using the `eig` command.)

You should find that only two eigenvalues are large, and therefore that the data can be well represented using two components only. Let $\mathbf{e}^1$ and $\mathbf{e}^2$ be the two eigenvectors with largest eigenvalues.

- Calculate the two dimensional representation of each datapoint $(\mathbf{e}^1 \cdot (\mathbf{x}^\mu - \mathbf{c}), \mathbf{e}^2 \cdot (\mathbf{x}^\mu - \mathbf{c})), \mu = 1, \ldots, 6$.
- Calculate the reconstruction of each datapoint $\mathbf{c} + (\mathbf{e}^1 \cdot (\mathbf{x}^\mu - \mathbf{c}))\mathbf{e}^1 + (\mathbf{e}^2 \cdot (\mathbf{x}^\mu - \mathbf{c}))\mathbf{e}^2$, $\mu = 1, \ldots, 6$.

3. Consider a set of N-dimensional data $\mathbf{x}^\mu, \mu = 1, \ldots, P$. A PCA analysis is performed on this dataset, and the eigenvectors $\mathbf{e}^1, \ldots, \mathbf{e}^M$ are used to represent the data, together with the mean $\mathbf{c}$. That is, each original N dimensional datapoint $\mathbf{x}^\mu$ is represented in the form

$$\mathbf{x}^\mu \approx \mathbf{c} + \sum_{i=1}^{M} a_i \mathbf{e}^i \tag{1}$$

- What are the optimal coefficients a_i if we are to minimise the square length of the residual vector?
- Consider two vectors $\mathbf{x}^a$ and $\mathbf{x}^b$ and their corresponding PCA approximations $\mathbf{c} + \sum_{i=1}^{M} a_i \mathbf{e}^i$ and $\mathbf{c} + \sum_{i=1}^{M} b_i \mathbf{e}^i$. Approximate $(\mathbf{x}^a - \mathbf{x}^b)^2$ by using the PCA representations of the data, and show that this is equal to $(\mathbf{a} - \mathbf{b})^2$.

4 **Additional Question.** Explain how the above result enables us to rapidly compare distances between PCA representations of datapoints, and how therefore in the Eigenfaces experiment, we can perform nearest neighbour classification rapidly, once the 20 largest eigenvectors have been calculated.