Advanced Data Analysis
Homework Week - 12

Aswin Vijay

Question 5: PCA derivation using maximum variance fromulation

Consider the dataset of observations $\{x_i\}$, $i = 1, ..., n$ and has dimension d. We then define an orthogonal basis $\left\{t_j | t_j \in \mathbb{R}^d\right\}_{j=1}^{m}$, where $m \leq d$ such that,

$$t_j^T t_{j'} = \begin{cases} 1 & (j = j') \\ 0 & (j \neq j') \end{cases} \tag{1}$$
$$TT^T = I, T = (t_i, ..., t_m)^T \in \mathbb{R}^{m \times d}$$

.

The orthogonal projection of sample x_i on this basis is then given by,

$$\sum_{j=1}^{m} (t_j^T x_i) \cdot t_j = T^T T x_i \tag{2}$$

Let the sample mean of the observations be,

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3}$$

The mean of the projected data is then

$$\sum_{j=1}^{m} (t_j^T \overline{x}) \cdot t_j = T^T T \overline{x} \tag{4}$$

The variance of the projeted data is then given by,

$$\frac{1}{n} \sum_{i=1}^{n} \left(T^T T x_i - T^T T \overline{x}\right)^2 \tag{5}$$

Let us define the covariance matrix as,

$$C = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})^T \tag{6}$$

Then the projected variance becomes,

$$\frac{1}{n} \sum_{i=1}^{n} \left(T^T T x_i - T^T T \overline{x}\right)^2 = T^T C T \tag{7}$$

We now maximise the projected variance $T^T C T$ with respect to T. This has to be a constrained maximisation to prevent $\|T\| \to \infty$. The constraint comes fromt the normalisation condition

$TT^T = I$. We introduce lagranges multiplier λ_1 to enforce the condition, the maximization objective then becomes,

$$\begin{gathered}\max_{T \in R^{m \times d}} T^T C T + \lambda_1 (I - TT^T) \\ T_{\text{PCA}} = \text{argmax}_{T \in R^{m \times d}} \ \text{tr}(T^T C T) \text{ s.t } TT^T = I_m\end{gathered} \tag{8}$$

Which is equivalent to the objective formulated by minimizing sum of projected errors. Setting derivative with respect to T equal to zero we get,

$$CT = \lambda_1 T \tag{9}$$

multiplying LHS by T^T we get,

$$T^T C T = \lambda_1 \tag{10}$$

So the variance will be a maximum when we set T equal to the eigenvector having the largest eigenvalue λ_1. The solution to the PCA objective is then obtained by eigen value decomposition of the C matrix and is given by,

$$T_{\text{PCA}} = U(\varepsilon_1, ..., \varepsilon_m)^T \tag{11}$$

- $\varepsilon_1, ..., \varepsilon_m$: The eigenvectors corresponding to the eigenvalues $\lambda_1 \geq \lambda_2 \geq ...\lambda_d$ of the eigenvalue problem $C\varepsilon = \lambda\varepsilon$.
- U: any $m \times m$ orthogonal matrix, where $U^{-1} = U^T$