CO-496: Mathematics for Inference and Machine Learning

Problem Sheet for Tutorial 4

Problem 1
Assume a matrix $\mathbf{X} \in \Re^{F \times n}$ where $n < F$. Prove that $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$ have the same positive eigenvalues. Furthermore, assume the eigen-decomposition of $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ and of $\mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$. Prove that the following holds

$$\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{\Lambda}^{-1/2}. \tag{1}$$

Problem 2
Assume the data samples $\mathbf{x}_1, \ldots, \mathbf{x}_n$, split in $C = 2$ different classes, with N_{C_1} and N_{C_2} number of samples, respectively. To deal with the drawback of LDA that finds only one dimension in two-class problems, a dimensionality reduction method called Biased-Discriminant Analysis (BDA) was proposed. The intuition behind BDA is that we aim to find a dimensionality reduction method that finds low dimensional features that minimise the variance of the first class, while the distance of the samples of the second class with regards to the centre of the first class is maximised. BDA finds the optimal $\mathbf{W}$ by solving the following trace optimisation problem

$$\begin{aligned} &\min_{\mathbf{W}} \ \mathrm{tr}(\mathbf{W}^T\mathbf{S}_1\mathbf{W}) \\ &\text{subject to } \ \mathbf{W}^T\mathbf{S}_2\mathbf{W} = \mathbf{I} \end{aligned} \tag{2}$$

where $\mathbf{S}_1 = \sum_{\mathbf{x}_i \in C_1}(\mathbf{x}_i - \mathbf{m}_{C_1})(\mathbf{x}_i - \mathbf{m}_{C_1})^T$ and $\mathbf{S}_2 = \sum_{\mathbf{x}_i \in C_2}(\mathbf{x}_i - \mathbf{m}_{C_1})(\mathbf{x}_i - \mathbf{m}_{C_1})^T$. Solve the above problem in Small Sampled Size problems.

Problem 3
A variation of BDA is the following. Find the optimal $\mathbf{W}$ by solving the trace optimisation problem

$$\begin{aligned} &\min_{\mathbf{W}} \ \mathrm{tr}(\mathbf{W}^T\mathbf{S}_1\mathbf{W}) - \mathrm{tr}(\mathbf{W}^T\mathbf{S}_2\mathbf{W}) \\ &\text{subject to } \ \mathbf{W}^T\mathbf{W} = \mathbf{I} \end{aligned} \tag{3}$$

Find the optimal $\mathbf{W}$ in Small Sampled Size problems, assuming that it can be written as a linear combination of the centralised samples.