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 $\mathbf{x}_1, \dots, \mathbf{x}_n$, 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

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

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

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

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

Solutions

Problem 1
Assume $\mathbf{u}$ is an eigenvector of $\mathbf{X}\mathbf{X}^T$, corresponding to an eigenvalue $\lambda > 0$. Therefore, it holds that $\mathbf{X}\mathbf{X}^T\mathbf{u} = \lambda\mathbf{u}$. We can thus have

$$\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{u}) = \lambda(\mathbf{X}^T\mathbf{u})$$

Since $\lambda \neq 0$, it follows that $\mathbf{X}^T\mathbf{u} \neq \mathbf{0}$, hence $\mathbf{X}^T\mathbf{u}$ is an eigenvector of $\mathbf{X}^T\mathbf{X}$ with the same positive eigenvalue λ. As a result, $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$ have the same positive eigenvalues and subsequently they have the following eigen-decomposition $\mathbf{X}\mathbf{X}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ and $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$. The second part is thoroughly proven in the notes.

Problem 2
We have $\mathbf{S}_2 = \mathbf{X}_2\mathbf{X}_2^T$, where $\mathbf{X}_2$ has as columns the vectors $\mathbf{x}_i - \mathbf{m}_{C_1}$ for all $\mathbf{x}_i$ of the second-class. We also have $\mathbf{S}_1 = \mathbf{X}_1\mathbf{X}_1^T$, where $\mathbf{X}_1$ has as columns the vectors $\mathbf{x}_i - \mathbf{m}_{C_1}$ for all $\mathbf{x}_i$ of the first-class. Assume that the optimal $\mathbf{W}$ is given by $\mathbf{W} = \mathbf{U}\mathbf{Q}$. Then the first step is to diagonalise $\mathbf{S}_2$ as

$$\mathbf{U}^T\mathbf{S}_2\mathbf{U} = \mathbf{I}. \tag{4}$$

Since $\mathbf{S}_2 = \mathbf{X}_2\mathbf{X}_2^T$, we will firstly need to perform eigen-decomposition on

$$\mathbf{X}_2^T\mathbf{X}_2 = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T. \tag{5}$$

We will then have $\mathbf{U} = \mathbf{X}_2\mathbf{V}\mathbf{\Lambda}^{-1}$. Furthermore, in order to find the optimal $\mathbf{Q}$, we have to solve

$$\begin{aligned} &\min_{\mathbf{Q}} \ \operatorname{tr}(\mathbf{Q}^T\mathbf{U}^T\mathbf{S}_1\mathbf{U}\mathbf{Q}) \\ &\text{subject to } \ \mathbf{Q}^T\mathbf{Q} = \mathbf{I}. \end{aligned} \tag{6}$$

As a result, $\mathbf{Q}$ contains as columns the d eigenvectors that correspond to the d largest eigenvalues of $\mathbf{U}^T\mathbf{S}_1\mathbf{U}$.

Problem 3
Assume $\mathbf{W} = \mathbf{X}\mathbf{Q}$. The optimisation problem can be then reformulated as

$$\begin{aligned} &\min_{\mathbf{Q}} \ \operatorname{tr}(\mathbf{Q}^T\mathbf{X}^T\mathbf{S}_1\mathbf{X}\mathbf{Q}) - \operatorname{tr}(\mathbf{Q}^T\mathbf{X}^T\mathbf{S}_2\mathbf{X}\mathbf{Q}) \\ &\text{subject to } \ \mathbf{Q}^T\mathbf{X}^T\mathbf{X}\mathbf{Q} = \mathbf{I} \end{aligned} \tag{7}$$

We now have to find $\mathbf{Q}$, where $\mathbf{Q} = \mathbf{R}\mathbf{G}$, where $\mathbf{R}$ is given by

$$\mathbf{R}^T\mathbf{X}^T\mathbf{X}\mathbf{R} = \mathbf{I}. \tag{8}$$

Assuming the eigen-decomposition of $\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$, then $\mathbf{R} = \mathbf{V}\mathbf{\Lambda}^{-1/2}$. The optimisation problem is now reformulated as

$$\begin{aligned} \min_{\mathbf{G}} \ & \mathrm{tr}(\mathbf{G}^T\mathbf{R}^T\mathbf{X}^T\mathbf{S}_1\mathbf{X}\mathbf{R}\mathbf{G}) - \mathrm{tr}(\mathbf{G}^T\mathbf{R}^T\mathbf{X}^T\mathbf{S}_2\mathbf{X}\mathbf{R}\mathbf{G}) \\ & \text{subject to } \ \mathbf{G}^T\mathbf{G} = \mathbf{I} \end{aligned} \tag{9}$$

The optimal $\mathbf{G}$ can be found by choosing the d eigenvectors that correspond to the d smallest eigenvalues of $\mathbf{R}^T\mathbf{X}^T(\mathbf{S}_1 - \mathbf{S}_2)\mathbf{X}\mathbf{R}$.

Another solution would be to observe that since we want to diagonalise $\mathbf{X}^T\mathbf{X}$, we just need to choose as original $\mathbf{W}$ a linear combination of the whitening transformation $\mathbf{U} = \mathbf{X}\mathbf{V}\mathbf{\Lambda}^{-1}$, i.e., $\mathbf{W} = \mathbf{X}\mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{G}$ (still a linear combination of $\mathbf{X}$). Then $\mathbf{G}$ can be found by solving

$$\begin{aligned} \min_{\mathbf{G}} \ & \mathrm{tr}(\mathbf{G}^T\mathbf{U}^T\mathbf{S}_1\mathbf{U}\mathbf{G}) - \mathrm{tr}(\mathbf{G}^T\mathbf{U}^T\mathbf{S}_2\mathbf{U}\mathbf{G}) \\ & \text{subject to } \ \mathbf{G}^T\mathbf{G} = \mathbf{I}. \end{aligned} \tag{10}$$

The optimal $\mathbf{G}$ can be found by choosing the d eigenvectors that correspond to the d smallest eigenvalues of $\mathbf{U}^T(\mathbf{S}_1 - \mathbf{S}_2)\mathbf{U}$.