D5: Databases and Information Systems
Data Mining and Matrices, SS 2017
Homework #6: ICA and spectral methods
Tutorial: **19 July 2017** at 10:15

You can discuss these problems with other students, but everybody must do and present their own answers. You can use computers etc. to perform the algebraic operations, but you must show the intermediate steps (and "computer said so" is never a valid answer). You are of course free to use material from the Internet, but again, you must present the intermediate steps and you must also be able to explain why the steps are valid and why you chose them. You can mark an answer even if it is not complete or correct, as long as you have made significant progress towards solving it. Note, however, that the TA does the final decision on whether your solution is complete (or correct) enough for a mark. ***We continue to apply more strict evaluation on what constitutes as sufficiently solved problem. You should only mark a problem if you think you have essentially solved it. This is done to give you a better impression on how the exam questions will be graded.***

Problem 1 (Kurtosis of a sum). Recall that the *kurtosis* of a random variable X with zero mean is

$$\mathrm{kurt}(X) = \mathbb{E}[X^4] - 3(\mathbb{E}[X^2])^2 \; . \tag{1.1}$$

One way to understand the importance of the factor 3 in (1.1) is to consider a sum of two independent random variables. Let X and Y be two independent random variables with zero mean and unit variance, i.e.

$$\mathbb{E}[X] = 0 \qquad \mathbb{E}[X^2] = 1 \tag{1.2}$$
$$\mathbb{E}[Y] = 0 \qquad \mathbb{E}[Y^2] = 1 \; . \tag{1.3}$$

Show that

$$\mathrm{kurt}(X + Y) = \mathrm{kurt}(X) + \mathrm{kurt}(Y) \; . \tag{1.4}$$

Can you see the importance of factor 3?

Hint: Use binomial formula and linearity of expectation.

Problem 2 (Kurtosis of normal distribution). Another way to see the importance of the factor 3 is to consider the kurtosis of normal distribution. We will prove that if X is normally distributed with 0 mean, then $\mathrm{kurt}(X) = 0$. To compute the kurtosis, we need the fourth moment $\mathbb{E}[X^4]$. To compute it, we use very powerful and general technique of *moment-generating functions*. The moment-generating function of random variable Y is

$$M_Y(t) = \mathbb{E}[\exp(tX)] = \mathbb{E}[\mathrm{e}^{tX}] \; , \quad t \in \mathbb{R} \; . \tag{2.1}$$

One important feature of moment-generating functions is that if we know M_Y, we can easily compute the nth moment of Y by differentiating M_Y n times and evaluating the derivative at origin. In other words,

$$\frac{d^n M_Y}{dt^n}(0) = \mathbb{E}[Y^n] \; , \tag{2.2}$$

where $\frac{d^n M_Y}{dt^n}(0)$ is the nth derivative of M_Y evaluated at origin. (Here we assume that the derivative exists.)

The moment-generating function for normally distributed X with 0 mean and variance σ^2 is

$$M_X(t) = \exp(\sigma^2 t^2/2) \; . \tag{2.3}$$

Use (2.3) to compute $\mathbb{E}[X^4]$ and conclude that $\mathrm{kurt}(X) = 0$.

Problem 3 (Traces and eigenvalues). Let $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \in \mathbb{R}^{n \times n}$ be a matrix and its eigendecomposition. You can assume that $\boldsymbol{Q}$ is orthogonal and that the eigenvalues are real. Show that

$$\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i = \mathrm{tr}(\boldsymbol{\Lambda}) \; . \tag{3.1}$$

Problem 4 (Laplacian is positive semi-definite). Let $G = (V, E)$ be an undirected graph and let $\boldsymbol{A}$ be its adjacency matrix and $\boldsymbol{\Delta}$ its degree matrix. Let $\boldsymbol{L} = \boldsymbol{\Delta} - \boldsymbol{A}$ be the Laplacian of G. Recall that the *incidence matrix* $\boldsymbol{P}$ of G (for some fixed but arbitrary ordering of the edges) is the $|V|$-by-$|E|$ matrix with

$$p_{ij} = \begin{cases} 1 & \text{if edge } j \text{ starts from node } i \\ -1 & \text{if edge } j \text{ ends to node } i \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

Show that

$$\boldsymbol{L} = \boldsymbol{P}\boldsymbol{P}^T \tag{4.2}$$

and conclude that the Laplacian is positive semi-definite.

Problem 5 (Normalized cut). Show that the solution for the relaxed normalized cut is obtained by taking the k least eigenvectors of the symmetric normalized Laplacian L^s similarly as the ratio cut is solved by taking the k least eigenvectors of the Laplacian.

Hint: Express normalized cut using the symmetric Laplacian by re-writing the equation

$$J_{nc}(\mathcal{C}) = \sum_{i=1}^{k} \frac{\boldsymbol{c}_i^T \boldsymbol{L} \boldsymbol{c}_i}{\boldsymbol{c}_i^T \boldsymbol{\Delta} \boldsymbol{c}_i}$$

using the facts that $\boldsymbol{\Delta} = \boldsymbol{\Delta}^{1/2}\boldsymbol{\Delta}^{1/2}$, $\boldsymbol{\Delta}^{1/2}\boldsymbol{\Delta}^{-1/2} = I$, and $\boldsymbol{\Delta} = \boldsymbol{\Delta}^T$ (as $\boldsymbol{\Delta}$ is diagonal).

Problem 6 (Nuclear norm). Let $\boldsymbol{A} \in \mathbb{R}^{n \times m}$ be an arbitrary matrix, and let $(\sigma_i)_{i=1}^{\min\{m,n\}}$ be its singular values. The *nuclear norm* of $\boldsymbol{A}$, denoted $\|\boldsymbol{A}\|_*$, is defined as

$$\|\boldsymbol{A}\| = \sum_{i=1}^{\min\{m,n\}} \sigma_i \ . \tag{6.1}$$

Show that

$$\|\boldsymbol{A}\|_* \geq \|\boldsymbol{A}\|_F \ . \tag{6.2}$$