Tsinghua-Berkeley Shenzhen Institute
LEARNING FROM DATA
Fall 2024

Writing Assignment 4

Issued: Friday 6th December, 2024 **Due:** Friday 20th December, 2024

POLICIES

- **Acknowledgments:** We expect you to make an honest effort to solve the problems individually. As we sometimes reuse problem set questions from previous years, covered by papers and web pages, we expect the students **NOT** to copy, refer to, or look at the solutions in preparing their answers (relating to an unauthorized material is considered a violation of the honor principle). Similarly, we expect you to not google directly for answers (though you are free to google for knowledge about the topic). If you do happen to use other material, it must be acknowledged here, with a citation on the submitted solution.
- **Required homework submission format:** You can submit homework either as one single PDF document or as handwritten papers. Written homework needs to be provided during the class on the due date, and a PDF document needs to be submitted through Tsinghua's Web Learning (`http://learn.tsinghua.edu.cn/`) before the end of the due date.
- **Collaborators:** In a separate section (before your answers), list the names of all people you collaborated with and for which question(s). If you did the HW entirely on your own, **PLEASE STATE THIS**. Each student must understand, write, and hand in answers of their own.

4.1. (SVD properties)

(a) (1 point) Prove that the left singular vectors of $\boldsymbol{A}$ are the right singular vectors of $\boldsymbol{A}^\top$.

(b) (2 points) If $\sigma_1, \sigma_2, \cdots, \sigma_r$ are the singular values of matrix $\boldsymbol{A}$ with rank r, and $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots, \boldsymbol{v}_r$ are the corresponding right singular vectors, show that: 1) $\boldsymbol{A}^\top \boldsymbol{A} = \sum_{i=1}^r \sigma_i^2 \boldsymbol{v}_i \boldsymbol{v}_i^\top$; 2) $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots, \boldsymbol{v}_r$ are eigenvectors of $\boldsymbol{A}^\top \boldsymbol{A}$.

(c) (2 points) For vector $c \in \mathbb{R}^n$, $d \in \mathbb{R}^m$ and matrix $\Omega \in \mathbb{R}^{n \times m}$. If $||c||^2 = ||d||^2 = 1$, prove that $c^\top \Omega d$ is maximized when c and d are the left and right unit singular vectors of Ω with the largest singular value.

(d) (Bonus 2 points) Prove that: 1) any matrix $\boldsymbol{M}$ with rank 1 can be written as the outer product of two vectors; 2) the sum of rank 1 matrices $\sum_{i=1}^r \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top$ can be written as $\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^\top$, where $\boldsymbol{u}_i$ are the columns of $\boldsymbol{U}$ and $\boldsymbol{v}_i$ are the columns of $\boldsymbol{V}$. *Hint: first verify that for any two matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, we have*

$$\boldsymbol{P}\boldsymbol{Q}^\top = \sum_i \boldsymbol{p}_i \boldsymbol{q}_i^\top, \tag{1}$$

where $\boldsymbol{p}_i$ is the i^{th} column of $\boldsymbol{P}$ and $\boldsymbol{q}_i$ is the i^{th} column of $\boldsymbol{Q}$.

4.2. (PCA) We look at another interpretation of PCA based on Projection Residual Minimization. Suppose we have m samples $\{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(m)}\}$, $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$, then we try to use the projections or image vectors to represent the original data. There will exist errors between the true data and their representations (projection residuals) and naturally we hope to minimize such errors.

(a) (2 points) First consider the case with one-dimensional projections. Let $\boldsymbol{u}$ be a non-zero unit vector. The projection of sample $\boldsymbol{x}^{(i)}$ on vector $\boldsymbol{x}u$ is represented by $(\boldsymbol{x}^{(i)\top}\boldsymbol{u})\boldsymbol{u}$. Therefore the residual of a projection will be

$$||\boldsymbol{x}^{(i)} - (\boldsymbol{x}^{(i)\top}\boldsymbol{u})\boldsymbol{u}||. \tag{2}$$

Please show that

$$\arg\min_{\boldsymbol{u}:\boldsymbol{u}^\top\boldsymbol{u}=1} ||\boldsymbol{x}^{(i)} - (\boldsymbol{x}^{(i)\top}\boldsymbol{u})\boldsymbol{u}||^2 = \arg\max_{\boldsymbol{u}:\boldsymbol{u}^\top\boldsymbol{u}=1}(\boldsymbol{x}^{(i)\top}\boldsymbol{u})^2. \tag{3}$$

(b) (2 points) Follow the proof above and the discussion of the variance of projections in the lecture. Please show that minimizing the residual of projections is equivalent to finding the largest eigenvector of covariance matrix Σ, *i.e.*, denote

$$\boldsymbol{u}^* = \arg\min_{\boldsymbol{u}:\boldsymbol{u}^\top\boldsymbol{u}=1} \frac{1}{m}\sum_{i=1}^{m} ||\boldsymbol{x}^{(i)} - (\boldsymbol{x}^{(i)\top}\boldsymbol{u})\boldsymbol{u}||^2, \tag{4}$$

then $\boldsymbol{u}^*$ is the largest eigenvector of $\boldsymbol{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}^{(i)}\boldsymbol{x}^{(i)\top}$.

4.3. (ICA) In the lecture, we briefly discussed why Gaussian random variables are forbidden in ICA. To understand this limitation more formally, let's assume that the joint distribution of two independent components, say, s_1, s_2, are Gaussian, *i.e.*,

$$P(s_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s_i^2}{2}\right) \tag{10}$$

(a) (1 point) Please find the joint pdf $P(s_1, s_2)$.

(b) (2 points) Suppose that the mixing matrix $\boldsymbol{A}$ is orthogonal. For example, we could assume that this is so because the data has been whitened, which means $\boldsymbol{A}^{-1} = \boldsymbol{A}^\top$ holds. Please find the joint pdf $P(x_1, x_2)$ of the mixtures x_1 and x_2 and then explain why Gaussian variables are forbidden.