Mathematics for Data Science [*]

Homework Assignment

M1 IDD, 2023-2024

Dauphine UNIVERSITÉ PARIS | PSL

Assignment

- This homework consists of proving several statements used without proofs in various papers accepted to machine learning conferences and journals.
- Students may discuss the project with their classmates, but answers must be written and submitted individually.
- Students may submit their answers in either French or English.
- Students are expected to send their answers in PDF format to `clement.royer@lamsade.dauphine.fr`. The PDF name must clearly indicate the student's first and last name.
- The deadline is set to **February 11, 2024 AOE** (Anywhere On Earth).

Major updates to the document

- 2024.01.22: Clarified the notation $\|\boldsymbol{x}\|$ (Euclidean norm) in Statement 3. Added a missing minus sign on the right-hand side of Statement 4.

[*] Version of January 22, 2024. The last version of this document can be found at: **https://www.lamsade.dauphine.fr/~croyer/ensdocs/MDS/DMMDS.pdf**

Statement 1 (Journal of Machine Learning Research, 2011)
Let $p \in [0,1]$, $\boldsymbol{A}, \boldsymbol{B} \succ \boldsymbol{0}$ *and* $\alpha \in [0,1]$. *Then,*

$$(\alpha \boldsymbol{A} + (1-\alpha)\boldsymbol{B})^p \succeq \alpha \boldsymbol{A}^p + (1-\alpha)\boldsymbol{B}^p$$

where two matrices $\boldsymbol{X}, \boldsymbol{Y} \succ \boldsymbol{0}$ *satisfy* $\boldsymbol{X} \succeq \boldsymbol{Y}$ *if and only if* $\boldsymbol{X} - \boldsymbol{Y} \succeq \boldsymbol{0}$.

Statement 2 (International Conference on Machine Learning, 2016)
Let $\boldsymbol{x}, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$ *be vectors in* $\mathbb{R}^d$ *such that* $\{\boldsymbol{v}_i\}_{i=1}^d$ *is an orthonormal basis of* $\mathbb{R}^d$ *and the variables* $\{\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{x}\}_{i=1}^d$ *are i.i.d. standard normal variables. Then,*

$$\mathbb{P}\left(\sqrt{\sum_{2\le i\le d} (\boldsymbol{v}_i^{\mathrm{T}}\boldsymbol{x})^2} \le \mathcal{C}d\right) \ge 1 - \exp\left(-\tfrac{d}{\mathcal{K}}\right),$$

where $\mathcal{C} > 0$ *and* $\mathcal{K} > 0$ *do not depend on* d.

Statement 3 (Conference on Learning Theory, 2022)
Let $\boldsymbol{v} \in \mathbb{R}^d$ *be a random vector uniformly distributed in* $\left\{-\frac{1}{\sqrt{d}}, \frac{1}{\sqrt{d}}\right\}^d$. *Then, for any vector* $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}$ *is a mean zero, sub-gaussian random variable such that*

$$\|\boldsymbol{x}^{\mathrm{T}}\boldsymbol{v}\|_{\psi_2} \le 2c\,\frac{\|\boldsymbol{x}\|}{\sqrt{d}}$$

where $\|\boldsymbol{x}\| = \sqrt{\sum_{i=1}^d x_i^2}$ *denotes the Euclidean norm, and* $c > 0$ *is a universal constant that does not depend on* d.

Statement 4 (Neural Information Processing Systems, 2023)
Consider a random matrix $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ *such that* $\|\boldsymbol{A}\|_F^2 = \sum_{j=1}^d \sum_{k=1}^d [\boldsymbol{A}]_{jk}^2$ *follows a chi-square distribution with degree of freedom* d^2, *and a value* $\epsilon > 0$. *Then,*

$$\mathbb{P}\left(\|\boldsymbol{A}\|_F^2 - d^2 \ge d^2 + \frac{3}{\epsilon}\right) \le \exp\left(-\frac{1}{\epsilon}\right).$$