Homework 3 - High Dimensional Probability

Bernardi Alberto, Cracco Gianmarco,
Manente Alessandro, Mazzieri Riccardo

December 2020

Exercise 1

Consider an $m \times n$ matrix A. Prove that:

$$\max_{x \in \mathcal{S}^{n-1}} \|Ax\|_2 = \max_{\substack{x \in \mathcal{S}^{n-1} \\ y \in \mathcal{S}^{m-1}}} \langle Ax, y \rangle.$$

Solution:

$$\langle Ax, y \rangle \leq \|Ax\|_2 \cdot \|y\|_2 = \|Ax\|_2 \leq \max_{x \in \mathcal{S}^{n-1}} \|Ax\|_2 \quad \forall y \in \mathcal{S}^{m-1}$$

$$\implies \max_{\substack{x \in \mathcal{S}^{n-1} \\ y \in \mathcal{S}^{m-1}}} \langle Ax, y \rangle \leq \max_{x \in \mathcal{S}^{n-1}} \|Ax\|_2.$$

Taking $y = \frac{Ax}{\|Ax\|_2} \in \mathcal{S}^{m-1}$ we got:

$$\langle Ax, \frac{Ax}{\|Ax\|_2} \rangle = \frac{\|Ax\|_2^2}{\|Ax\|_2} = \|Ax\|_2$$

$$\implies \max_{\substack{x \in \mathcal{S}^{n-1} \\ y \in \mathcal{S}^{m-1}}} \langle Ax, y \rangle = \max_{x \in \mathcal{S}^{n-1}} \|Ax\|_2.$$

$\blacksquare$

Exercise 2

Let A be a real matrix with $\text{rank}(A) = r$. Let us indicate with $\|A\|_F$ and $\|A\|$ the Frobenius and operator norm respectively.

(i) Prove that $\|A\|_F^2 = \text{Tr}(A^\top A)$. Deduce that $\|A\|_F = \left(\sum_{i=1}^r s_i^2(A)\right)^{1/2}$.

(ii) Prove that $\|A\| \leq \|A\|_F \leq \sqrt{r} \cdot \|A\|$.

Solution:

(i) Given A an $m \times n$ matrix, recall that $\|A\|_F^2 = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2$.
We then have:

$$\operatorname{Tr}(A^\top A) = \sum_{j=1}^n (A^\top A)_{ij} = \sum_{j=1}^n \left(a_{1j}^2 + \ldots + a_{mj}^2\right) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = \|A\|_F^2.$$

Moreover, recalling that $s_i(A) = \sqrt{\lambda_i(A^\top A)}$ we have:

$$\|A\|_F = \sqrt{\operatorname{Tr}(A^\top A)} = \sqrt{\sum_{i=1}^r \lambda_i(A^\top A)} = \sqrt{\sum_{i=1}^r s_i^2(A)} = \left(\sum_{i=1}^r s_i^2(A)\right)^{1/2}.$$

(ii) Recall that $\|A\| = s_1(A)$, where $s_1(A)$ is the maximum singular value of A, and notice that:

$$\|A\| = \sqrt{s_1^2(A)} \leq \sqrt{\sum_{i=1}^r s_i^2(A)} = \sqrt{\sum_{i=1}^r \lambda_i(A^\top A)} = \sqrt{\operatorname{Tr}(A^\top A)} = \|A\|_F.$$

Furthermore:

$$\|A\|_F = \sqrt{\sum_{i=1}^r \lambda_i(A^\top A)} \leq \sqrt{\sum_{i=1}^r \lambda_1(A^\top A)} = \sqrt{r \cdot s_1^2(A)} = \sqrt{r} \cdot \|A\|.$$

■

Exercise 3

Prove that for any s_i singular value of A it holds:

$$s_i \leq \frac{1}{\sqrt{i}} \|A\|_F.$$

Solution:

We know that $\|A\|_F = \left(\sum_{i=1}^n s_i^2(A)\right)^{1/2}$ so:

$$\|A\|_F \geq \left(\sum_{j=1}^i s_i^2(A) + \sum_{k=i+1}^n s_k^2(A)\right)^{1/2} \geq \left(i \cdot s_i^2(A)\right)^{1/2} = \sqrt{i} \cdot s_i(A)$$

$$\implies s_i \leq \frac{1}{\sqrt{i}} \|A\|_F.$$

■

Exercise 4

Prove that the Hamming distance d_H is a metric.

Solution:

- $d_H(X,Y) = 0 \iff X = Y \quad \forall X, Y \in \{0,1\}^n$:

$$d_H(X,Y) = 0 \implies \nexists i : X(i) \neq Y(i) \implies X = Y$$

$$X = Y \implies \nexists i : X(i) \neq Y(i) \implies d_H(X,Y) = 0.$$

- $d_H(X,Y) = d_H(Y,X)$:

  follows from symmetry of $\neq$.

- $d_H(X,Y) \leq d_H(X,Z) + d_H(Z,Y)$:

  let us define the following: $A_{XZ} = \{i | X(i) \neq Z(i)\}$, $A_{ZY} = \{i | Z(i) \neq Y(i)\}$ and $A_{XY} = \{i | X(i) \neq Y(i)\}$, which are all subsets of $\{1, \dots, n\}$.
  Now consider $i \in A_{XY}$, then we have two cases:

  - $i \in A_{XZ} \implies i \notin A_{ZY}$;
  - $i \in A_{ZY} \implies i \notin A_{XZ}$.

  Therefore we have $A_{XY} = (A_{XZ} \setminus A_{ZY}) \cup (A_{ZY} \setminus A_{XZ})$ and so:

$$\begin{aligned} d_H(X,Y) = |A_{XY}| &= |(A_{XZ} \setminus A_{ZY}) \cup (A_{ZY} \setminus A_{XZ})| \\ &= |A_{XZ} \setminus A_{ZY}| + |A_{ZY} \setminus A_{XZ}| \leq |A_{XZ}| + |A_{ZY}| \\ &= d_H(X,Z) + d_H(Z,Y). \end{aligned}$$

■

Exercise 5

Let $K = \{0,1\}^n$. Prove that for every integer $m \in [0,n]$ we have

$$\frac{2^n}{\sum_{k=0}^{m} \binom{n}{k}} \overset{(1)}{\leq} \mathcal{N}(K, d_H, m) \overset{(2)}{\leq} \mathcal{P}(K, d_H, m) \overset{(3)}{\leq} \frac{2^n}{\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \binom{n}{k}}.$$

Solution:

Let us study the inequalities separately:

(1) $\frac{2^n}{\sum_{k=0}^{m} \binom{n}{k}} \leq \mathcal{N}(K, d_H, m)$:

We first recall that $\mathcal{N}$ can be seen as the minimum number of closed balls of radius m that cover K. We now find how many elements of K are

contained in a ball of radius m. Calling $X_0 \in K$ the center of the ball, $Y \in B_{X_0,m}$ if $d_H(X_0, Y) \leq m \iff |\{i \in \mathbb{N} | X(i) \neq Y(i)\}| \leq m$ (i.e. $B_{X_0,m}$ contains all $Y \in K$ such that Y differs from X_0 in at most m entries).
The number of such Y is therefore: $|B_m| = \binom{n}{0} + \ldots + \binom{n}{m} = \sum_{k=0}^{m} \binom{n}{k}$.
Hence:

$$\mathcal{N}(K, d_H, m) \cdot |B_m| \geq |K| \iff \frac{2^n}{\sum_{k=0}^{m} \binom{n}{k}} \leq \mathcal{N}(K, d_H, m)$$

(2) $\mathcal{N}(K, d_H, m) \leq \mathcal{P}(K, d_H, m)$:

trivial from the theory.

(3) $\mathcal{P}(K, d_H, m) \leq \frac{2^n}{\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \binom{n}{k}}$:

Also here we recall that $\mathcal{P}$ can be seen as the maximum number of disjoint balls of radius $\frac{m}{2}$ centred in K. Notice that, considering any $B_{X_0, \frac{m}{2}} = \{x | d_H(x, X_0) \leq \frac{m}{2}\}$, we have $B_{X_0, \frac{m}{2}} \subset K \quad \forall X_0 \in K$. This holds true because $d_H(x, y) \leq n \quad \forall x, y \in K$. From an intuitive point of view this means that such balls cannot "go out" of K. Furthermore, since such balls are disjoint, it holds true that their union has cardinality less or equal to the cardinality of K. More formally:

$$\mathcal{P}(K, d_H, m) \cdot |B_{\frac{m}{2}}| \leq |K|$$
$$\Updownarrow$$
$$\mathcal{P}(K, d_H, m) \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \binom{n}{k} \leq 2^n$$
$$\Updownarrow$$
$$\mathcal{P}(K, d_H, m) \leq \frac{2^n}{\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \binom{n}{k}}.$$

$\blacksquare$

Exercise 6

Let A be an $m \times n$ matrix, $\mu \in \mathbb{R}$ and $\varepsilon \in [0, 1/2)$. Show that for any ε-net $\mathcal{N}$ of the sphere $\mathcal{S}^{n-1}$ we have:

$$\sup_{x \in \mathcal{S}^{n-1}} \left| \|Ax\|_2 - \mu \right| \leq \frac{C}{1 - 2\varepsilon} \cdot \sup_{x \in \mathcal{N}} |\ \|Ax\|_2 - \mu\ |\ .$$

Solution:
First, note that we can assume, without loss of generality, that $\mu = 1$: indeed, dividing both terms of the target inequality by the same quantity μ, it is clear that the result does not change. We can now rewrite $\|Ax\|^2$ as follows:

$$\|Ax\|^2 = \langle Ax, Ax\rangle = (Ax)^\top Ax = x^\top A^\top Ax = \langle A^\top Ax, x\rangle.$$

Now, since $x \in \mathcal{S}^{n-1}$, we have that $\|x\|_2^2 = 1 = \langle x, x\rangle$. Therefore:

$$\|Ax\|_2^2 - 1 = \langle A^\top Ax, x\rangle - \langle x, x\rangle = \langle \underbrace{(A^\top A - \mathbf{1}_n)}_{R} x, x\rangle = \langle Rx, x\rangle.$$

Note that R is a $n \times n$ symmetric matrix. Now, recalling the fact that $|z-1| \leq |z^2 - 1| \quad \forall z \geq 0$, we can write:

$$\sup_{x\in\mathcal{S}^{n-1}} |\|Ax\|_2 - 1| \leq \sup_{x\in\mathcal{S}^{n-1}} |\|Ax\|_2^2 - 1| = \sup_{x\in\mathcal{S}^{n-1}} |\langle Rx, x\rangle| = \dots$$

Now, since $\max_{x\in\mathcal{S}^{n-1}} \|Ax\|_2 = \max_{x\in\mathcal{S}^{n-1}} \langle Ax, x\rangle$ and observing that the supremum and the maximum coincide being $\mathcal{S}^{n-1}$ a compact set, we can write:

$$\dots = \sup_{x\in\mathcal{S}^{n-1}} \|Rx\|_2 = \|R\| \leq \frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\langle Rx, x\rangle|,$$

where we used the result from Exercise 4.4.3[1] in the last inequality. To conclude:

$$\begin{aligned}
\frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\langle Rx, x\rangle| &\leq \frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\|Ax\|_2^2 - 1| \\
&= \frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |(\|Ax\|_2 - 1)\underbrace{(\|Ax\|_2 + 1)}_{\leq 1+\|A\|=C}| \\
&\leq \frac{C}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\|Ax\|_2 - 1|.
\end{aligned}$$

■

[1]Let A be a $n \times n$ symmetric square matrix and $\varepsilon \in [0, 1/2)$. For any ε-net $\mathcal{N}$ of the sphere $\mathcal{S}^{n-1}$ we have

$$\sup_{x\in\mathcal{N}} |\langle Ax, x\rangle| \leq \|A\| \leq \frac{1}{1-2\varepsilon} \sup_{x\in\mathcal{N}} |\langle Ax, x\rangle|.$$

Exercise 7

Deduce from Theorem 4.4.5[2] that:

$$\mathbb{E}\|A\| \leq C\left(\sqrt{m}+\sqrt{n}\right).$$

Solution:
From Thm. 4.4.5, $\forall t > 0$, $\|A\| \leq CK(\sqrt{m}+\sqrt{n}+t)$ with probability at least $1-2\exp(-t^2)$

$$\implies \mathbb{P}\{\|A\| > CK(\sqrt{m}+\sqrt{n}+t)\} \leq 2\exp\left(-t^2\right)$$

$$\underbrace{\int_{-\infty}^{+\infty} \mathbb{P}\{\|A\| > CK(\sqrt{m}+\sqrt{n}+t)\}dt}_{y=CK(\sqrt{m}+\sqrt{n}+t)\Rightarrow dt=\frac{1}{CK}dy} \leq \int_{-\infty}^{+\infty} 2\exp\left(-t^2\right)dt$$

$$\int_{-\infty}^{+\infty} \mathbb{P}\{\|A\| > y\}\cdot\frac{1}{CK}dy = \frac{1}{CK}\mathbb{E}\|A\| \leq 2\sqrt{\pi}$$

$$\begin{aligned}\implies \mathbb{E}\|A\| &\leq CK\cdot 2\sqrt{\pi}\\ &= CK\\ &\leq CK(\sqrt{m}+\sqrt{n})\end{aligned}$$

■

[2] Let A be an $m\times\, n$ random matrix whose entries A_{ij} are independent, mean zero, subgaussian random variables. Then, for any $t>0$ we have:

$$\|A\| \leq CK(\sqrt{m}+\sqrt{n}+t)$$

with probability at least $1-2\exp\left(-t^2\right)$. Here $K = \max_{i,j}\left\|A_{ij}\right\|_{\psi_2}$ and C is a positive constant.