

# **Exercise of Supervised Learning: Curse of Dimensionality**

Yawei Li

`yawei.li@stat.uni-muenchen.de`

November 26, 2024

# Exercise 1

Consider a random vector  $X = (X_1, \dots, X_p)^T \sim \mathcal{N}(0, I)$ , i.e., a multivariate normally distributed vector with mean vector zero and covariance matrix being the identity matrix of dimension  $p \times p$ . In this case, the coordinates  $X_1, \dots, X_p$  are i.i.d. each with distribution  $\mathcal{N}(0, 1)$ .

(a) Show that  $\mathbb{E}[\|X\|_2^2] = p$  and  $\text{Var}(\|X\|_2^2) = 2p$ , where  $\|\cdot\|_2$  is the Euclidean norm.

Hint:  $\mathbb{E}_{Y \sim \mathcal{N}(0,1)}(Y^4) = 3$ .

# Solution to Exercise 1 (a)

Note that

$$||X||_2^2 = \sum_{i=1}^p X_i^2.$$

Then,

$$\mathbb{E}[||X||_2^2] = \mathbb{E}\left[\sum_{i=1}^p X_i^2\right]$$

# Solution to Exercise 1 (a)

Note that

$$\|X\|_2^2 = \sum_{i=1}^p X_i^2.$$

Then,

$$\begin{aligned}\mathbb{E}[\|X\|_2^2] &= \mathbb{E}\left[\sum_{i=1}^p X_i^2\right] \\ &= \sum_{i=1}^p \mathbb{E}[X_i^2]\end{aligned}$$

# Solution to Exercise 1 (a)

Note that

$$||X||_2^2 = \sum_{i=1}^p X_i^2.$$

Then,

$$\begin{aligned}\mathbb{E}[||X||_2^2] &= \mathbb{E}\left[\sum_{i=1}^p X_i^2\right] \\ &= \sum_{i=1}^p \mathbb{E}[X_i^2] \\ &= \sum_{i=1}^p (\underbrace{\mathbb{E}[X_i]^2}_{=0} + \text{Var}(X_i))\end{aligned}$$

# Solution to Exercise 1 (a)

Note that

$$||X||_2^2 = \sum_{i=1}^p X_i^2.$$

Then,

$$\begin{aligned}\mathbb{E}[||X||_2^2] &= \mathbb{E}\left[\sum_{i=1}^p X_i^2\right] \\ &= \sum_{i=1}^p \mathbb{E}[X_i^2] \\ &= \sum_{i=1}^p (\underbrace{\mathbb{E}[X_i]^2}_{=0} + \text{Var}(X_i)) \\ &= \sum_{i=1}^p 1\end{aligned}$$

# Solution to Exercise 1 (a)

Note that

$$||X||_2^2 = \sum_{i=1}^p X_i^2.$$

Then,

$$\begin{aligned}\mathbb{E}[||X||_2^2] &= \mathbb{E}\left[\sum_{i=1}^p X_i^2\right] \\ &= \sum_{i=1}^p \mathbb{E}[X_i^2] \\ &= \sum_{i=1}^p (\underbrace{\mathbb{E}[X_i]^2}_{=0} + \text{Var}(X_i)) \\ &= \sum_{i=1}^p 1 \\ &= p.\end{aligned}$$

## Solution to Question 1 (a): Continued

$$\begin{aligned}\mathrm{Var}(\|X\|_2^2) &= \mathrm{Var}\left(\sum_{i=1}^p X_i^2\right) \\&= \sum_{i=1}^p \mathrm{Var}(X_i^2) \\&= \sum_{i=1}^p \left(\underbrace{\mathbb{E}[X_i^4]}_{=3} - \underbrace{\mathbb{E}[X_i^2]^2}_{=1}\right) \quad \triangleright \\&= \sum_{i=1}^p (3 - 1) \\&= 2p.\end{aligned}$$



## Exercise 1 (b)

(b) Use (a) to infer that  $|\mathbb{E}[\|X\|_2 - \sqrt{p}]| \leq \frac{1}{\sqrt{p}}$  by using the following steps:

(i) Write  $\|X\|_2 - \sqrt{p} = \underbrace{\frac{\|X\|_2 - p}{2\sqrt{p}}}_{:= (1)} - \underbrace{\frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2}}_{:= (2)}.$

(ii) Compute  $\mathbb{E}[(1)]$ .

(iii) Note that  $0 \leq \mathbb{E}[(2)] \leq \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}}$  holds due to  $\|X\|_2 \geq 0$ .

(iv) Put (i)- (iii) together.

## Solution to Exercise 1 (b)

Step (i): Write  $\|X\|_2 - \sqrt{p} = \underbrace{\frac{\|X\|_2 - p}{2\sqrt{p}}}_{:= (1)} - \underbrace{\frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2}}_{:= (2)}.$   $\triangleleft$

Step (ii):

$$\begin{aligned}\mathbb{E}[(1)] &= \mathbb{E}\left[\frac{\|X\|_2^2 - p}{2\sqrt{p}}\right] \\ &= \frac{1}{2\sqrt{p}}\left(\underbrace{\mathbb{E}[\|X\|_2^2]}_{=p \text{ (from Question(a))}} - p\right) \\ &= 0.\end{aligned}$$

## Solution to Exercise 1 (b)

Step (i): Write  $\|X\|_2 - \sqrt{p} = \underbrace{\frac{\|X\|_2 - p}{2\sqrt{p}}}_{:= (1)} - \underbrace{\frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2}}_{:= (2)}.$   $\triangleleft$

Step (ii):

$$\begin{aligned}\mathbb{E}[(1)] &= \mathbb{E}\left[\frac{\|X\|_2^2 - p}{2\sqrt{p}}\right] \\ &= \frac{1}{2\sqrt{p}}\left(\underbrace{\mathbb{E}[\|X\|_2^2]}_{=p \text{ (from Question(a))}} - p\right) \\ &= 0.\end{aligned}$$

## Solution to Exercise 1(b): Continued

Step (iii): Prove that  $0 \leq \mathbb{E}[(2)] \leq \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}}$  holds due to  $\|X\|_2 \geq 0$ .

$$\mathbb{E}[(2)] = \mathbb{E} \left[ \frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2} \right] \geq 0, \quad \text{since all the terms are non-negative.}$$

Besides, since  $\|X\|_2 \geq 0$ , it follows that

$$\begin{aligned} (2) &\leq \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \\ \Rightarrow \mathbb{E}[(2)] &\leq \mathbb{E} \left[ \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \right] = \frac{1}{2p^{3/2}} \cdot \mathbb{E} [(\|X\|_2^2 - \mathbb{E}[\|X\|_2^2])^2] = \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}} \\ &= \frac{2p}{2p^{3/2}} = \frac{1}{\sqrt{p}}, \end{aligned}$$

where we utilize the lemma from (a) that  $\mathbb{E}[\|X\|_2^2] = p$  and  $\text{Var}(\|X\|_2^2) = 2p$ .

## Solution to Exercise 1(b): Continued

Step (iii): Prove that  $0 \leq \mathbb{E}[(2)] \leq \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}}$  holds due to  $\|X\|_2 \geq 0$ .

$$\mathbb{E}[(2)] = \mathbb{E} \left[ \frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2} \right] \geq 0, \quad \text{since all the terms are non-negative.}$$

Besides, since  $\|X\|_2 \geq 0$ , it follows that

$$\begin{aligned} (2) &\leq \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \\ \Rightarrow \mathbb{E}[(2)] &\leq \mathbb{E} \left[ \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \right] = \frac{1}{2p^{3/2}} \cdot \mathbb{E} [(\|X\|_2^2 - \mathbb{E}[\|X\|_2^2])^2] = \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}} \\ &= \frac{2p}{2p^{3/2}} = \frac{1}{\sqrt{p}}, \end{aligned}$$

where we utilize the lemma from (a) that  $\mathbb{E}[\|X\|_2^2] = p$  and  $\text{Var}(\|X\|_2^2) = 2p$ .

## Solution to Exercise 1(b): Continued

Step (iii): Prove that  $0 \leq \mathbb{E}[(2)] \leq \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}}$  holds due to  $\|X\|_2 \geq 0$ .

$$\mathbb{E}[(2)] = \mathbb{E} \left[ \frac{(\|X\|_2^2 - p)^2}{2\sqrt{p}(\|X\|_2 + \sqrt{p})^2} \right] \geq 0, \quad \text{since all the terms are non-negative.}$$

Besides, since  $\|X\|_2 \geq 0$ , it follows that

$$\begin{aligned} (2) &\leq \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \\ \Rightarrow \mathbb{E}[(2)] &\leq \mathbb{E} \left[ \frac{(\|X\|_2^2 - p)^2}{2p^{3/2}} \right] = \frac{1}{2p^{3/2}} \cdot \mathbb{E} [(\|X\|_2^2 - \mathbb{E}[\|X\|_2^2])^2] = \frac{\text{Var}(\|X\|_2^2)}{2p^{3/2}} \\ &= \frac{2p}{2p^{3/2}} = \frac{1}{\sqrt{p}}, \end{aligned}$$

where we utilize the lemma from (a) that  $\mathbb{E}[\|X\|_2^2] = p$  and  $\text{Var}(\|X\|_2^2) = 2p$ .

## Solution to Exercise 1(b): Continued

Step (iv): Putting everything together:

$$|\mathbb{E}[||X||_2 - \sqrt{p}]| = |\underbrace{\mathbb{E}[(1)]}_{=0} - \underbrace{\mathbb{E}[(2)]}_{\geq 0}| = \mathbb{E}[(2)] \leq \frac{1}{\sqrt{p}}.$$

## Exercise 1 (c)

(c) Use (b) to infer that  $\text{Var}(\|X\|_2) \leq 2$  by using the following steps:

- (i) Write  $\text{Var}(\|X\|_2) = \text{Var}(\|X\|_2 - \sqrt{p})$ .
- (ii) For any random variable  $Y$  it holds that  $\text{Var}(Y) \leq \mathbb{E}[Y^2]$ .
- (iii) If you encounter the term  $E[\|X\|_2]$  write it as  $\mathbb{E}[\underbrace{\|X\|_2 - \sqrt{p}}_{= (*)} + \sqrt{p}]$  and use (b) for  $(*)$ .



## Solution to Exercise 1 (c)

Step (i): Write  $\text{Var}(\|X\|_2) = \text{Var}(\|X\|_2 - \sqrt{p})$ .

It holds because **variance does not change by constant shifts**.

Step (ii): For any random variable  $Y$  it holds that  $\text{Var}(Y) \leq \mathbb{E}[Y^2]$ .

It holds because  $\text{Var}(Y) + \mathbb{E}[Y]^2 = \mathbb{E}[Y^2]$  and  $\mathbb{E}[Y]^2 \geq 0$ . Later we will use this inequality.

## Solution to Exercise 1 (c): Continued

$$\begin{aligned}\text{Var}(\|X\|_2) &= \text{Var}(\|X\|_2 - \sqrt{p}) && \text{Step (i)} \\&\leq \mathbb{E}[(\|X\|_2 - \sqrt{p})^2] && \text{Step (ii)} \\&= \mathbb{E}[\|X\|_2^2 - 2\sqrt{p}\|X\|_2 + p] \\&= \underbrace{\mathbb{E}[\|X\|_2^2]}_{=p} - 2\sqrt{p} \cdot \mathbb{E}[\|X\|_2] + p \\&= 2p - 2\sqrt{p} \cdot \mathbb{E}[\|X\|_2] \\&= 2p - 2\sqrt{p} \cdot \mathbb{E}[\|X\|_2 - \sqrt{p} + \sqrt{p}] && \text{Step (iv)} \\&= 2p - 2p - 2\sqrt{p} \cdot \mathbb{E}[\|X\|_2 - \sqrt{p}] \\&= -2\sqrt{p} \cdot \underbrace{\mathbb{E}[\|X\|_2 - \sqrt{p}]}_{\in [-\frac{1}{\sqrt{p}}, \frac{1}{\sqrt{p}}]} && \text{(from (b))} \\&\leq 2\sqrt{p} \cdot \frac{1}{\sqrt{p}} = 2.\end{aligned}$$

## Question 1 (d)

Now let  $X' = (X'_1, \dots, X'_p)^T \sim \mathcal{N}(0, \mathbf{I})$  be another multivariate normally distributed vector with mean vector zero and covariance matrix being the identity matrix of dimension  $p \times p$ . Further, assume that  $X$  and  $X'$  are independent, so that  $Z := \frac{X - X'}{\sqrt{2}} \sim \mathcal{N}(0, \mathbf{I})$ . Conclude from the previous that

$$\left| \mathbb{E} \left[ \|X - X'\|_2 - \sqrt{2p} \right] \right| \leq \sqrt{\frac{2}{p}} \quad \text{and} \quad \text{Var}(\|X - X'\|_2) \leq 4.$$

## Solution to Exercise 1 (d)

We first investigate  $Z$ . Since  $Z = \frac{X-X'}{\sqrt{2}} \sim \mathcal{N}(0, I)$ , it follows from (b) and (c) that

$$|\mathbb{E}[||Z||_2 - \sqrt{p}]| \leq \sqrt{\frac{1}{p}}, \quad (1)$$

$$\text{Var}(|Z|_2) \leq 2 \quad (2)$$

But the norm of  $Z$

$$||Z||_2 = \sqrt{\sum_{i=1}^p \left(\frac{X_i - X'_i}{\sqrt{2}}\right)^2} = \sqrt{\frac{1}{2} \sum_{i=1}^p (X_i - X'_i)^2} = \sqrt{\frac{1}{2}} \sqrt{\sum_{i=1}^p (X_i - X'_i)^2} = \sqrt{\frac{1}{2}} ||X - X'||_2. \quad (3)$$

It follows from (1) that

$$\sqrt{2} \cdot |\mathbb{E}[||Z||_2 - \sqrt{p}]| \leq \sqrt{\frac{2}{p}} \Rightarrow |\mathbb{E}[\underbrace{\sqrt{2}||Z||_2}_{||X-X'||_2} - \sqrt{2p}]| \leq \sqrt{\frac{2}{p}}.$$

## Solution to Exercise 1 (d)

We first investigate  $Z$ . Since  $Z = \frac{X-X'}{\sqrt{2}} \sim \mathcal{N}(0, I)$ , it follows from (b) and (c) that

$$|\mathbb{E}[\|Z\|_2 - \sqrt{p}]| \leq \sqrt{\frac{1}{p}}, \quad (1)$$

$$\text{Var}(\|Z\|_2) \leq 2 \quad (2)$$

But the norm of  $Z$

$$\|Z\|_2 = \sqrt{\sum_{i=1}^p \left(\frac{X_i - X'_i}{\sqrt{2}}\right)^2} = \sqrt{\frac{1}{2} \sum_{i=1}^p (X_i - X'_i)^2} = \sqrt{\frac{1}{2}} \sqrt{\sum_{i=1}^p (X_i - X'_i)^2} = \sqrt{\frac{1}{2}} \|X - X'\|_2. \quad (3)$$

It follows from (1) that

$$\sqrt{2} \cdot |\mathbb{E}[\|Z\|_2 - \sqrt{p}]| \leq \sqrt{\frac{2}{p}} \Rightarrow \underbrace{|\mathbb{E}[\sqrt{2}\|Z\|_2 - \sqrt{2p}]|}_{\|X - X'\|_2} \leq \sqrt{\frac{2}{p}}.$$

## Solution to Exercise 1 (d)

We first investigate  $Z$ . Since  $Z = \frac{X-X'}{\sqrt{2}} \sim \mathcal{N}(0, I)$ , it follows from (b) and (c) that

$$|\mathbb{E}[\|Z\|_2 - \sqrt{p}]| \leq \sqrt{\frac{1}{p}}, \quad (1)$$

$$\text{Var}(\|Z\|_2) \leq 2 \quad (2)$$

But the norm of  $Z$

$$\|Z\|_2 = \sqrt{\sum_{i=1}^p \left(\frac{X_i - X'_i}{\sqrt{2}}\right)^2} = \sqrt{\frac{1}{2} \sum_{i=1}^p (X_i - X'_i)^2} = \sqrt{\frac{1}{2}} \sqrt{\sum_{i=1}^p (X_i - X'_i)^2} = \sqrt{\frac{1}{2}} \|X - X'\|_2. \quad (3)$$

It follows from (1) that

$$\sqrt{2} \cdot |\mathbb{E}[\|Z\|_2 - \sqrt{p}]| \leq \sqrt{\frac{2}{p}} \Rightarrow \underbrace{|\mathbb{E}[\sqrt{2}\|Z\|_2 - \sqrt{2p}]|}_{\|X - X'\|_2} \leq \sqrt{\frac{2}{p}}.$$

## Solution to Exercise 1 (d): Continued

Moreover, (2):  $\text{Var}(\|Z\|_2) \leq 2$  implies that

$$\text{Var}(\|Z\|_2) \leq 2$$

$$\Leftrightarrow 2\text{Var}(\|Z\|_2) \leq 4$$

$$\Leftrightarrow \text{Var}(\sqrt{2}\|Z\|_2) \leq 4 \quad (\text{Var}(aY) = a^2\text{Var}(Y) \text{ for any RV } Y)$$

$$\Leftrightarrow \text{Var}(\|X - X'\|_2) \leq 4 \quad \text{Using (3) that } \|Z\|_2 = \sqrt{\frac{1}{2}}\|X - X'\|_2.$$

## Exercise 1 (e)

(e) From the cosine rule we can infer that for any  $x, x' \in \mathbb{R}^p$  it holds that

$$\langle x, x' \rangle = \frac{1}{2}(\|x\|_2^2 + \|x'\|_2^2 - \|x - x'\|_2^2).$$

Use this to show that  $\mathbb{E}[\langle X, X' \rangle] = 0$ . Moreover, derive that  $\text{Var}(\langle X, X' \rangle) = p$ .



## Solution to Exercise 1 (e)

Since  $\langle x, x' \rangle = \frac{1}{2}(\|x\|_2^2 + \|x'\|_2^2 - \|x - x'\|_2^2)$ , we can infer that

$$\begin{aligned}\mathbb{E}[\langle X, X' \rangle] &= \frac{1}{2}(\mathbb{E}[\|X\|_2^2] + \mathbb{E}[\|X'\|_2^2] - \mathbb{E}[\|X - X'\|_2^2]) \\ &= \frac{1}{2} \left( p + p - 2 \cdot \mathbb{E} \left[ \underbrace{\frac{1}{2} \|X - X'\|_2^2}_{=\|Z\|_2^2} \right] \right) \\ &= \frac{1}{2}(p + p - 2p) = 0. \quad (\text{From (a) we know that } \mathbb{E}[\|Z\|_2^2] = p)\end{aligned}$$

## Solution to Exercise 1 (e): Continued

$$\begin{aligned}\text{Var}(\langle X, X' \rangle) &= \text{Var}\left(\sum_{i=1}^p X_i X'_i\right) \\&= \sum_{i=1}^p \text{Var}(X_i X'_i) \\&= p \text{Var}(X_1 X'_1) \\&= p \cdot (\mathbb{E}[X_1^2 (X'_1)^2] - \mathbb{E}[X_1 (X'_1)]^2) \\&= p \cdot \underbrace{(\mathbb{E}[X_1^2])}_{=1} \cdot \underbrace{\mathbb{E}[(X'_1)^2]}_{=1} - \underbrace{\mathbb{E}[X_1]^2}_{=0} \cdot \underbrace{\mathbb{E}[X'_1]^2}_{=0} \\&= p.\end{aligned}$$

## Exercise 1 (f)

(f) For different dimensions  $p$ , e.g.,  $p \in \{1, 2, 4, 8, \dots, 1024\}$ , create two sets consisting of 100 i.i.d. random observations from  $\mathcal{N}(0, \mathbf{I})$ , respectively and

- (i) compute the average Euclidean length of (one of) the sampled sets and compare it to  $\sqrt{p}$ ;
- (ii) compute the average Euclidean distances between the sampled sets and compare it to  $\sqrt{2p}$ ;
- (iii) compute the average inner products between the sampled sets;
- (iv) compute in (i)-(iii) also the empirical variances of the respective terms.

Visualize your results in an appropriate manner.

**Show the standard solution.**