

**This homework is due at 11 PM on February 13, 2026.**

**Submission Format:** Your homework submission should consist of a single PDF file that contains all of your answers (any handwritten answers should be scanned).

**1. Norms**

- (a) Show that the following inequalities hold for any vector  $\vec{x} \in \mathbb{R}^n$ :

$$\frac{1}{\sqrt{n}} \|\vec{x}\|_2 \leq \|\vec{x}\|_\infty \leq \|\vec{x}\|_2 \leq \|\vec{x}\|_1 \leq \sqrt{n} \|\vec{x}\|_2 \leq n \|\vec{x}\|_\infty. \quad (1)$$

*NOTE:* We can interpret different norms as different ways of computing distance between two points  $\vec{x}, \vec{y} \in \mathbb{R}^2$ . The  $\ell^2$  norm is the distance as the crow flies (i.e. point-to-point distance), the  $\ell^1$  norm, also known as the Manhattan distance is the distance you would have to cover if you were to navigate from  $\vec{x}$  to  $\vec{y}$  via a rectangular street grid, and the  $\ell^\infty$  norm is the maximum distance that you have to travel in either the north-south or the east-west direction.

- (b) We define the *sparsity* of the vector  $\vec{x}$  as the number of non-zero elements in  $\vec{x}$ . This is also commonly known as the  $\ell^0$  norm of the vector  $\vec{x}$ , denoted by  $\|\vec{x}\|_0$ . Show that for any non-zero vector  $x$ ,

$$\|\vec{x}\|_0 \geq \frac{\|\vec{x}\|_1^2}{\|\vec{x}\|_2^2}. \quad (2)$$

Find all vectors  $\vec{x}$  for which the lower bound is attained.

**2. Diagonalization and Singular Value Decomposition**

Let matrix  $A = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ .

- (a) Compute the eigenvalues and associated eigenvectors of  $A$ .
- (b) Express  $A$  as  $P\Lambda P^{-1}$ , where  $\Lambda$  is a diagonal matrix and  $PP^{-1} = I$ . State  $P$ ,  $\Lambda$ , and  $P^{-1}$  explicitly.
- (c) Compute  $\lim_{k \rightarrow \infty} A^k$ .
- (d) Give the singular values  $\sigma_1$  and  $\sigma_2$  of  $A$ .

### 3. Interpreting the Data Matrix

When working in many fields, you'll often find yourself working with a *data matrix*  $X$ . Notation can vary — sometimes it has dimensions  $\mathbb{R}^{m \times n}$ , while others it has dimensions  $\mathbb{R}^{n \times d}$ , for example — and interpreting its precise meaning can often be confusing. In this problem, we lead you through an example of data matrix interpretation and manipulation.

First, what exactly is a data matrix? As the name suggests, it is a collection of *data points*. Suppose you are collecting data about courses offered in the EECS department in Fall 2022. Each course has certain quantifiable attributes, or *features*, that you are interested in. Possible examples of features are the number of students in the course, the number of GSIs in the course, the number of units the course is worth, the size of the classroom that the course was taught in, the difficulty rating of the course on a numerical (1-5) scale, and so on. Suppose there were a total of 20 courses, and that for each course, we have 10 features. This gives us 20 data points, where each data point is a 10-dimensional vector. We can arrange these data points in a matrix of size  $20 \times 10$ .

Generalizing the above, suppose we have  $n$  data points, with each point containing values for  $d$  features. Our data matrix  $X$  would then be of size  $n \times d$ , i.e.,  $X \in \mathbb{R}^{n \times d}$ . We can interpret  $X$  in the following two (equivalent) ways:

$$X = \begin{bmatrix} \leftarrow \vec{x}_1^\top \rightarrow \\ \leftarrow \vec{x}_2^\top \rightarrow \\ \vdots \\ \leftarrow \vec{x}_n^\top \rightarrow \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \vec{f}_1 & \vec{f}_2 & \cdots & \vec{f}_d \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix}. \quad (3)$$

Here,  $\vec{x}_i \in \mathbb{R}^d$ ,  $i = 1, 2, \dots, n$ , and  $\vec{x}_i^\top$  is a row vector that contains values of different features for the  $i$ -th data point. Also,  $\vec{f}_j \in \mathbb{R}^n$ ,  $j = 1, 2, \dots, d$ , and  $\vec{f}_j$  is a column vector that contains values of the  $j$ -th feature for different data points.

In the remainder of this problem, we explore how we can interpret and use  $X$ . For subproblems that require answers in Python, assume  $X$  is stored as a  $n \times d$  NumPy array  $\mathbf{X}$ .

- (a) We first introduce the *empirical mean* of each feature. Let  $k \geq 1$  be a positive integer, and define  $\vec{1}$  to be the vector with 1 in every entry. The empirical mean of a vector  $\vec{y} \in \mathbb{R}^k$  is defined as

$$\mu(\vec{y}) \doteq \frac{1}{k} \vec{1}^\top \vec{y} = \frac{1}{k} \sum_{i=1}^k y_i. \quad (4)$$

Suppose we want to compute a vector that contains the empirical mean of each feature, i.e., all the  $\mu(\vec{f}_j)$ 's. What is the length of the vector of empirical means? Which of the following Python commands will generate this vector?

- i. `mu = numpy.mean(X, axis = 0)`
- ii. `mu = numpy.mean(X, axis = 1)`

- (b) The next quantity we will discuss is the *empirical variance*, and through it, the *empirical standard deviation*. The empirical variance of a vector  $\vec{y} \in \mathbb{R}^k$  is defined as

$$\sigma^2(\vec{y}) \doteq \frac{1}{k} \|\vec{y} - \mu(\vec{y})\vec{1}\|_2^2 = \frac{1}{k} \sum_{i=1}^k (y_i - \mu(\vec{y}))^2. \quad (5)$$

As the choice of notation would have you expect, the empirical standard deviation is defined as

$$\sigma(\vec{y}) \doteq \sqrt{\sigma^2(\vec{y})}. \quad (6)$$

Suppose we want to compute a vector that contains the empirical standard deviation of each feature, i.e., all the  $\sigma(\vec{f}_j)$ 's. What is the length of this vector? Which of the following Python commands will generate this vector?

- i. `sigma = numpy.std(X, axis = 0)`
  - ii. `sigma = numpy.std(X, axis = 1)`
- (c) Suppose we want to modify  $X$  so that each feature vector is “centered”, i.e., has zero empirical mean. How would you achieve this using Python code?
- (d) Suppose we want to modify  $X$  so that each feature vector is “standardized”, i.e., has zero empirical mean with empirical variance equal to 1. How would you achieve this using Python code?
- NOTE:* This standardization technique is a very common data pre-processing step.
- (e) The last quantity we will discuss is the *empirical covariance*. For two vectors  $\vec{w}, \vec{y} \in \mathbb{R}^k$ , the empirical covariance is defined as

$$\sigma(\vec{w}, \vec{y}) \doteq \frac{1}{k} (\vec{w} - \mu(\vec{w})\vec{1})^\top (\vec{y} - \mu(\vec{y})\vec{1}) = \frac{1}{k} \sum_{i=1}^k (w_i - \mu(\vec{w}))(y_i - \mu(\vec{y})). \quad (7)$$

What is  $\sigma(\vec{y}, \vec{y})$  in terms of the empirical statistics we have previously defined (e.g. mean, variance, and/or standard deviation)?

- (f) For the remainder of this problem, assume that the data matrix is centered, so every feature has zero empirical mean; that is,  $\mu(\vec{f}_j) = 0$  for every  $j$ .

Let  $\Sigma(X) \in \mathbb{R}^{d \times d}$  denote the *empirical covariance matrix* of  $X$ . This matrix contains the empirical covariance of each pair of feature vectors  $(\vec{f}_i, \vec{f}_j)$ . Correspondingly it is defined entry-wise as

$$\Sigma(X)_{i,j} \doteq \sigma(\vec{f}_i, \vec{f}_j). \quad (8)$$

First, show that

$$\Sigma(X) = \frac{X^\top X}{n}. \quad (9)$$

Second, show that

$$\frac{X^\top X}{n} = \frac{1}{n} \sum_{i=1}^n \vec{x}_i \vec{x}_i^\top. \quad (10)$$

Therefore, (??) entails that  $\Sigma(X)$  can be represented in two ways.

*HINT:* One way to show two matrices are equal is to show that for all  $i, j$ , their  $(i, j)$ -th entries are equal.

- (g) In this class, we consider three different interpretations of the term “projection”. We define them explicitly here for this problem.

Consider vectors  $\vec{a}$  and  $\vec{b}$  in  $\mathbb{R}^n$ . Let  $\vec{b}$  be unit norm (i.e.,  $\|\vec{b}\|_2^2 = \vec{b}^\top \vec{b} = 1$ ). We define the following:

- i. The **vector projection** of  $\vec{a}$  on  $\vec{b}$  is given by  $(\vec{a}^\top \vec{b})\vec{b}$ . The vector projection is a vector in  $\mathbb{R}^n$ .
- ii. The **scalar projection** of  $\vec{a}$  on  $\vec{b}$  is given by  $\vec{a}^\top \vec{b}$ . The scalar projection is a scalar but can take both positive and negative values.
- iii. The **projection length** of  $\vec{a}$  on  $\vec{b}$  is given by  $|\vec{a}^\top \vec{b}|$  and is the absolute value of the scalar projection.

Suppose we want to obtain a column vector  $\vec{p} \in \mathbb{R}^n$  whose  $i$ -th entry is the *scalar* projection of data point  $\vec{x}_i$  along the direction given by the unit vector  $\vec{w}$ . Show that  $\vec{p}$  is given by

$$\vec{p} = X\vec{w}. \quad (11)$$

- (h) Performing this kind of projection onto a unit vector  $\vec{w}$  is at the heart of the PCA computation, which also requires computing the *variance* of these scalar projections.

Formally, for  $i \in \{1, \dots, n\}$ , define  $p_i \doteq \vec{x}_i^\top \vec{w}$ , and define  $\vec{p} \doteq [p_1, \dots, p_n]^\top$ . Show that its empirical variance  $\sigma^2(\vec{p})$  can be calculated as

$$\sigma^2(\vec{p}) = \frac{1}{n} \vec{w}^\top X^\top X \vec{w} = \vec{w}^\top \Sigma(X) \vec{w}. \quad (12)$$

Recall that  $X$  is centered.

#### 4. Understanding Ellipses

Consider the Euclidean space  $\mathbb{R}^2$  with the orthogonal basis  $\{\vec{e}_1, \vec{e}_2\}$ . In this exercise, we study the ellipse

$$\mathcal{E} = \left\{ x_1 \vec{e}_1 + x_2 \vec{e}_2 \mid x_1, x_2 \in \mathbb{R}, \left( \sqrt{5}x_1 - \frac{3\sqrt{5}}{5}x_2 \right)^2 + \left( \frac{4\sqrt{5}}{5}x_2 \right)^2 \leq 8 \right\}. \quad (13)$$

- (a) Show that we can express the ellipse as  $\mathcal{E} = \{\vec{x} \in \mathbb{R}^2 \mid \vec{x}^\top A \vec{x} \leq 1\}$  for symmetric positive definite  $A$ , where

$$A = \frac{1}{8} \begin{bmatrix} 5 & -3 \\ -3 & 5 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}. \quad (14)$$

- (b) Show that the ellipse  $\mathcal{E}$  can be viewed as a linear transformation of the unit disk by finding  $B$  such that  $\mathcal{E} = \{B\vec{v} \mid \|\vec{v}\|_2 \leq 1\}$ . Is this  $B$  unique?

- (c) Relate the length and direction of the semi-major and semi-minor axes of  $\mathcal{E}$  to the singular values of  $B$  (or eigenvalues of  $A$ ).

- (d) Compute the area of  $\mathcal{E}$ .

**5. SVD Part 2**

Consider  $A$  to be the  $4 \times 3$  matrix

$$A = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix} \quad (15)$$

where  $\vec{a}_i$  for  $i \in \{1, 2, 3\}$  form a set of *orthogonal* vectors satisfying  $\|\vec{a}_1\|_2 = 3$ ,  $\|\vec{a}_2\|_2 = 2$ ,  $\|\vec{a}_3\|_2 = 1$ .

- (a) What is the compact SVD of  $A$ ? Express it as  $A = U\Sigma V^\top$ , with  $\Sigma$  the diagonal matrix of singular values ordered in decreasing fashion, and explicitly describe  $U$  and  $V$ .

- (b) What is the dimension of the null space,  $\dim(\mathcal{N}(A))$ ?

- (c) What is the rank of  $A$ ,  $\text{rank}(A)$ ? Provide an orthonormal basis for the range of  $A$ .

- (d) Let  $I_3$  denote the  $3 \times 3$  identity matrix. Consider the matrix  $\tilde{A} = \begin{bmatrix} A \\ I_3 \end{bmatrix} \in \mathbb{R}^{7 \times 3}$ . What are the singular values of  $\tilde{A}$  (in terms of the singular values of  $A$ )?

**6. Homework Process**

With whom did you work on this homework? List the names and SIDs of your group members.

*NOTE:* If you didn't work with anyone, you can put "none" as your answer.