Homework 2

Math 76 Summer 2026

Name: ________________________________

Problem 1. Let U be an orthogonal matrix.

(a) Show that $U^\mathsf{T} U = I$, where U^T is the transpose of U and I is the identity matrix.

(b) Show that if $x \in \mathbb{R}^n$ and $\|x\|^2 = x_1^2 + \cdots + x_n^2$, then

$$\|Ux\|^2 = \|x\|^2 .$$

Problem 2. Let A be an $m \times n$ matrix with $m \geq n$, and let $b \in \mathbb{R}^m$. The goal of this problem is to find $x \in \mathbb{R}^n$ that minimizes

$$\|Ax - b\|^2 .$$

Let $A = U\Sigma V^\mathsf{T}$ be a (full) singular value decomposition, where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal and $\Sigma \in \mathbb{R}^{m\times n}$ is diagonal with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

(a) Substitute the SVD for A to rewrite the objective:

$$\|Ax - b\|^2 \; = \; \|\underline{\hspace{3cm}}\|^2 .$$

(b) Using the fact that $\|Uy\| = \|y\|$ for an orthogonal U, factor U out of the norm and simplify:

$$\left\|U\Sigma V^\mathsf{T}x - b\right\|^2 \; = \; \|\underline{\hspace{5cm}}\|^2 .$$

(c) Apply the change of variables $y = V^\mathsf{T}x$ and $c = U^\mathsf{T}b$. What does the objective become in terms of y and c?

(d) Write $\|\Sigma y - c\|^2$ explicitly in terms of the σ_i, y_i, and c_i. (Remember that Σ is $m \times n$ with $m \geq n$, so the sum has "leftover" terms.)

(e) Find the optimal y_i, and then express the optimal x in matrix form. Use the pseudoinverse Σ^+, defined by $\Sigma^+_{ii} = 1/\sigma_i$ when $\sigma_i > 0$ (and 0 otherwise).

Problem 3. Discovering the SVD Autoencoder (guided).

Setup. You are given n data points $x_1, \dots, x_n \in \mathbb{R}^d$ that have already been *centered* (each feature has mean zero). Collect them as the columns of a matrix $X \in \mathbb{R}^{d \times n}$, and fix a target dimension $k < d$. Your goal is to *compress* each point to k numbers and then *reconstruct* it as accurately as possible, using only linear maps. Work through the parts in order; each builds on the last, taking you from the definition of a linear autoencoder to a clean maximization problem. (The next problem then solves that maximization with the SVD.)

(a) **Define the autoencoder.** An *encoder* is a linear map that compresses, and a *decoder* is a linear map that reconstructs. Write them as matrices: an encoder matrix E and a decoder matrix D, so the code is $z = Ex$ and the reconstruction is $\hat{x} = Dz = DEx$. Give the shapes of E and D, and show that the total squared error over the dataset is

$$\mathcal{L}(E, D) \;=\; \sum_{i=1}^{n} \|x_i - DEx_i\|^2 \;=\; \|X - DEX\|_F^2 .$$

What is the largest possible rank of $P = DE$, and why does this force some information to be lost when $k < d$?

(b) **Reduce to a rank constraint.** Show that as E, D range over the shapes in part (a), the product $M = DE$ ranges over *exactly* the $d \times d$ matrices with $\operatorname{rank}(M) \le k$. Conclude that the problem is equivalent to

$$\min_{\operatorname{rank}(M) \le k} \|X - MX\|_F^2 .$$

Hint. Use a rank factorization: any matrix of rank $\le k$ splits as a $(d \times k)$ times a $(k \times d)$ matrix.

(c) **Recognize the decoder as a projection.** Fix the subspace $S = \mathrm{col}(D)$ with $\dim S = k$. Among all $\hat{x} \in S$, which is closest to a given x? Use this to argue that we may assume D has orthonormal columns, with $E = D^\top$, so that $\hat{x} = DD^\top x$.

Hint. Projection theorem: $DD^\top$ is the orthogonal projector onto S exactly when $D^\top D = I_k$. The task is now to pick the best k-dimensional subspace.

(d) **Turn "smallest error" into a maximization.** Assume from now on that $D^\top D = I_k$, and write the reconstruction as $\hat{X} = DD^\top X$. First check that the residual is orthogonal to the reconstruction, $\mathrm{tr}\big(\hat{X}^\top(X - \hat{X})\big) = 0$, which gives the Pythagorean split

$$\|X\|_F^2 \;=\; \left\|\hat{X}\right\|_F^2 + \left\|X - \hat{X}\right\|_F^2 .$$

Next, using $D^\top D = I_k$, show that $\left\|\hat{X}\right\|_F^2 = \|D^\top X\|_F^2$. Since $\|X\|_F^2$ is a fixed constant, conclude that minimizing the error $\left\|X - \hat{X}\right\|_F^2$ is the *same* as maximizing

$$\left\|D^\top X\right\|_F^2 \;=\; \mathrm{tr}\big(D^\top X X^\top D\big),$$

the total squared length of the codes.

Problem 4. Solving the autoencoder with the SVD.

In the previous problem, finding the best rank-k linear autoencoder was reduced to the following optimization: among all $d \times k$ matrices D with orthonormal columns (that is, $D^\top D = I_k$), *maximize*

$$\left\| D^\top X \right\|_F^2 = \operatorname{tr}\left(D^\top X X^\top D\right),$$

where $X \in \mathbb{R}^{d\times n}$ collects the centered data as its columns and $k < d$. In this problem you solve this maximization using *only* the SVD (no eigenvalues or eigenvectors). Let $X = U\Sigma V^\top$ be a singular value decomposition, where $U \in \mathbb{R}^{d\times d}$ and $V \in \mathbb{R}^{n\times n}$ are orthogonal and $\Sigma \in \mathbb{R}^{d\times n}$ carries the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ on its main diagonal.

(a) Set $W = U^\top D$. Show that W still has orthonormal columns ($W^\top W = I_k$), and, using the fact that multiplying by the orthogonal matrices U and V leaves the Frobenius norm unchanged, show that

$$\left\| D^\top X \right\|_F^2 = \left\| W^\top \Sigma \right\|_F^2.$$

(b) Let r_i denote the squared norm of the i-th row of W. Show that

$$\left\| W^\top \Sigma \right\|_F^2 = \sum_i \sigma_i^2 \, r_i.$$

(c) Explain why $0 \leq r_i \leq 1$ for every i (the columns of W are orthonormal, so its rows cannot be too long), and why $\sum_i r_i = \|W\|_F^2 = k$.

(d) Since $\sigma_1^2 \geq \sigma_2^2 \geq \cdots$, conclude that $\sum_i \sigma_i^2 r_i$ is largest when $r_1 = \cdots = r_k = 1$ and the remaining $r_i = 0$, giving the maximum value $\sigma_1^2 + \cdots + \sigma_k^2$. Check that this is attained by $W = \left[\begin{smallmatrix} I_k \\ 0 \end{smallmatrix}\right]$, i.e. by taking D to be the first k columns of U. Hence the optimal autoencoder keeps the top k left singular vectors:

$$\boxed{E = U_k^\top, \qquad D = U_k.}$$