CIS3333: Mathematics of Machine Learning Fall 2025

Homework 5

Due: October 22, 2025, 11:59 PM ET

Submission Instructions: Submit a single PDF (clear scans or photos compiled) to Gradescope and assign pages for each problem. Show key steps and justify answers.

Collaboration & AI Policy: You may discuss approaches with classmates, but write up your own solutions and list collaborators. If you use computational tools (including LLMs) for checking, cite them and ensure the reasoning is your own.

Problem 1: Affine Subspaces and Solution Sets (7 points)

In class, we saw that the solution set to a homogeneous system of linear equations, $X\theta = 0$, is always a vector subspace (the kernel). This problem explores the structure of solutions for the more general non-homogeneous case, $X\theta = Y$.

Definition: An **affine subspace** L is a set formed by taking a vector subspace U and shifting it by a vector x_0. Formally,

$$L = x_0 + U = \{x_0 + u : u \in U\}$$

Geometrically, a line or plane that does not pass through the origin is a classic example of an affine subspace.

1. (3 points) An affine subspace is not always a vector subspace. Prove that the affine subspace $L = x_0 + U$ is a vector subspace if and only if $x_0 \in U$.
2. (3 points) Let θ_p be any particular solution to a consistent linear system $X\theta = Y$. Prove that the full solution set S can be written as the affine subspace $S = \theta_p + \ker(X)$.
3. (1 point) Using the structure of the solution set from above, state the specific condition on the dimension of the kernel, $\dim(\ker(X))$ such that the system has: (i) a unique solution or (ii) infinitely many solutions.

Problem 2: Uniqueness of Basis Size (10 points)

In this problem, we will prove this that the size of all bases for a vector space V are the same by showing that if there are two bases $\mathcal{B} = \{b_1, \ldots, b_n\}$ and $\mathcal{C} = \{c_1, \ldots, c_m\}$ for a vector space V, then $n = m$.

Our proof will be by contradiction. First, without loss of generality, we can assume $m \geq n$ to begin with. For contradiction, assume that $m > n$.

1. (2 points) Show that $\{c_1, b_1, \ldots, b_n\}$ is linearly dependent.

2. (2 points) Writing $c_1 = \sum_{i=1}^{n} \alpha_i b_i$, show that at least one of the scalar coefficients α_i must be non-zero. Let this coefficient be α_j.
3. (4 points) Use the above argument to show that $\mathcal{B}_1 = \{c_1\} \cup (\mathcal{B} \setminus \{b_j\})$ is a basis for V.
4. (2 points) Now we can repeat the same argument n times, yielding $\mathcal{B}_n = \{c_1, \ldots, c_n\}$, which is a basis for V. Explain how this contradicts that $\mathcal{C}$ is a basis.

Since we have reached a contradiction, we must have $m = n$.

Problem 3: Two Views of PCA (5 points)

The goal of PCA is to find a lower-dimensional representation of data that captures the most information. There are two perspectives on capturing the most information: **maximizing the variance** or **minimizing the reconstruction error** of the projected data.

Let the data be $x_1, \ldots, x_N \in \mathbb{R}^D$, and assume it has mean zero ($\frac{1}{N} \sum_i x_i = 0$). For simplicity, let us consider the case where we project the data down to a one-dimensional subspace. Let $\pi_u(x_i)$ be the orthogonal projection of x_i onto the span of u.

The variance maximization view of PCA suggests that we should find the direction u that maximizes the variance of the projected data. The variance of the projected data is given by:

$$\text{Var}(u) = \frac{1}{N} \sum_{i=1}^{N} \|\pi_u(x_i)\|^2.$$

An alternative way to solve this problem is to instead minimize the reconstruction error of the projected data. The reconstruction error is given by:

$$J(u) = \frac{1}{N} \sum_{i=1}^{N} \|x_i - \pi_u(x_i)\|^2$$

Show that minimizing $J(u)$ is equivalent to maximizing the variance of the projected data, that is,

$$\arg \min_{u, \|u\|=1} J(u) = \arg \max_{u, \|u\|=1} \text{Var}(u).$$

Problem 4: Similar Matrices (8 points)

Consider a matrix A. We say that A is similar to a matrix B if there exists an invertible matrix S such that $B = S^{-1}AS$.

1. (4 points) Show that the eigenvalues of A and B are the same. (Hint: You may need to use the properties of the determinant: $\det(XY) = \det(X)\det(Y)$ and $\det(X^{-1}) = 1/\det(X)$).
2. (2 points) Show that if v is an eigenvector of A with eigenvalue λ, then $S^{-1}v$ is an eigenvector of B with eigenvalue λ.
3. (2 points) Show that if A is diagonalizable, then B is also diagonalizable.