Homework 10
High-Dimensional Probability for Data Science, Fall 2023

Hints are in the back of the homework set.

One of the greatest advantages of PCA, compared to other dimension reduction and visualization methods, is that the results of PCA are easy to interpret. The first principal component explains the most variance, the second principal component explains the most remaining variance, etc. In this problem, we make this statement precise.

1. Percentage of explained variance by PCA

Let X be a random vector in $\mathbb{R}^d$ with mean zero and covariance matrix Σ. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ be the eigenvalues of Σ and $v_1, v_2, \ldots, v_d$ be the corresponding eigenvectors. Let P_k denote the orthogonal projection in $\mathbb{R}^d$ onto the span of $v_1, \ldots, v_k$.

(a) Show that

$$\mathbb{E}\|P_k X\|_2^2 = \sum_{i=1}^{k} \operatorname{Var}\left(\langle X, v_i \rangle\right) = \sum_{i=1}^{k} \lambda_i.$$

(b) Interpret the result of (a) as follows: the fraction of explained variance of the data equals by the first k principal components of PCA equals s_k/s_d, where $s_k = \sum_{i=1}^{k} \lambda_i$.

The two most widely used matrix norms are Frobenius and operator norms. For some matrices these two norms can return the same value, while for other matrices there can be a gap between the two norms, and the gap can be as large as the square root of the dimension. Let us prove these facts.

2. Operator vs.Frobenius norms of matrices

Let $m, n \in \mathbb{N}$.

(a) Prove that every $m \times n$ matrix A satisfies

$$\|A\| \leq \|A\|_F \leq \sqrt{r}\|A\|$$

where $r = \operatorname{rank}(A)$.

(b) Find an $m \times n$ matrix A for which $\|A\|_F = \|A\|$.

(c) For any value $1 \leq r \leq \min(m, n)$, find a matrix A with $\operatorname{rank}(A) = r$ and such that $\|A\|_F = \sqrt{r}\|A\|$.

In this problem we express the operator norm as the maximum of a quadratic form. This interpretation will be crucial in our analysis of covariance estimation.

3. Operator norm and quadratic forms

Let A be an $n \times n$ symmetric matrix.

(a) Show that

$$\|A\| = \max_{i=1,\dots,n} |\lambda_i(A)|\,,$$

where $\lambda_i(A)$ denote the eigenvalues of A.

(b) Show that

$$\|A\| = \max_{x \in S^{n-1}} \left| x^\mathsf{T} A x \right|,$$

where S^{n-1} denotes the unit Euclidean sphere, i.e. the set of unit vectors in $\mathbb{R}^n$.

(c) Show by example that the formula in (b) may fail for non-symmetric matrices.

In Problem 3 we showed how to compute the operator norm of A by maximizing the quadratic form $x^\mathsf{T} A x$ over the unit sphere. Now we show that the sphere can be replaced by an its ε-net. This obviously makes the maximum smaller. However, as we will see now, the net still captures the operator norm quite accurately. This discretization will be crucial in our analysis of covariance estimation.

4. Computing the operator norm on a net

Let A be an $n \times n$ symmetric matrix, $\varepsilon \in (0, 1/2)$, and $\mathcal{N}$ be an ε-net of the unit sphere S^{n-1}. Show that

$$\|A\| \le \frac{1}{1 - 2\varepsilon} \cdot \max_{x \in \mathcal{N}} \left| x^\mathsf{T} A x \right|.$$

TURN OVER FOR HINTS

Hints

Hints for Problem 1. Check that $P_k = \sum_{i=1}^k v_i v_i^\mathsf{T}$. Multiply this equation by X, compute the norm of the result using the orthogonality of vectors v_i, and finally take expectation. This should yield the first equation in the problem. To prove the second equation, check that $\mathrm{Var}\,(\langle X, v_i \rangle) = \lambda_i$ using the same trick as in Lecture 18: express $\langle X, v_i \rangle^2 = v_i^\mathsf{T} X X^\mathsf{T} v_i$ and take expectation on both sides.

Hints for Problem 2. (a) Consider a matrix whose all entries except one equal 0. (b) Consider a matrix whose r diagonal entries equal 1 and all other entries equal 0.

Hints for Problem 3. (a) Recall that the operator norm of A is the same as the maximal singular value of A. Since the matrix A is symmetric, its singular values equal $|\lambda_i(A)|$ (why?)

(b) Check that $\|Ax\|_2 \geq |x^\mathsf{T} Ax|$ for any unit vector x, and that an equality is attained for $x = u_1$, the top eigenvector of A.

(c) If A is a rotation of the plane by 90°, the vectors x and Ax are always orthogonal.

Hint for Problem 4. The overall approach can be similar to the proof of a similar result for non-symmetric matrices (Lecture 19, or Lemma 4.4.1 in the book). Choose a vector x that maximizes the quadratic form on the sphere, and approximate x with some vector u in the net. To bound the norm of the difference between $x^\mathsf{T} Ax$ and $u^\mathsf{T} Au$, bound the norm of the difference between each of these two quantities to $x^\mathsf{T} Au$ and use triangle inequality.