

Homework 9

Due Tuesday, November 24

Please either give the assignment to Loraine at the CDS or send it via email to the graders **before noon**.

1. *True or false (10 points)*. Are these statements true? Prove them or provide a counterexample.
 - a. The projection of a vector on a subspace \mathcal{S} is equal to

$$\mathcal{P}_{\mathcal{S}} x = \sum_{i=1}^n \langle x, b_i \rangle b_i$$

for any basis b_1, \dots, b_n of \mathcal{S} .

- b. The orthogonal complement of the orthogonal complement of a subspace \mathcal{S} is \mathcal{S} (assume that the ambient space has finite dimension).
 - c. If a matrix $A \in \mathbb{R}^n$ has $n \times n$ eigenvalues such that $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ corresponding to n linearly independent eigenvectors v_1, \dots, v_n , the power method converges to a vector in the span of $\{v_1\}$ no matter how you initialize it.
2. *Heartbeat (15 points)* We are interested in computing the best linear estimate of the heartbeat of a fetus in the presence of strong interference in the form of the heartbeat of the baby's mother. To simplify matters, let us assume that we only want to estimate the heartbeat at a certain moment. We have available a measurement from a microphone situated near the mother's belly and another from a microphone that is away from her belly. We model the measurements as

$$Y_1 = B + M \tag{1}$$

$$Y_2 = M + N. \tag{2}$$

B is a random variable modeling the heartbeat of the baby, M is a random variable modeling the heartbeat of the mother and N is a random variable modeling some additive noise. From past data, we determine that all random variables have zero mean and are uncorrelated. The variances of B and N are equal to σ^2 whereas the variance of M is much larger, it is equal to $3\sigma^2$.

- a. Apply Gram-Schmidt to compute a basis for the span of Y_1 and Y_2 within the vector space of zero-mean random variables.
 - b. Compute the best linear estimator of B given Y_1 and Y_2 in terms of MSE. Interpret the result.
3. *Cold (10 points)*. When Bob has a cold the probability that he is ill the day after is 0.8. When he is healthy, the probability that he is ill the day after is 0.1.
 - a. What is the probability that Bob has a cold today? State any assumptions you make.
 - b. If Bob has a cold today, estimate whether he had a cold three days before using MAP estimation.

4. *Wheat (15 points)*. The file *seeds_dataset.txt* contains different geometric attributes (area, perimeter, compactness, length of kernel, width of kernel, asymmetry coefficient and length of kernel groove) of seeds belonging to three different varieties of wheat: Kama, Rosa and Canadian¹. We want to visualize the data in 2D by projecting it, while capturing as much of its energy as possible.
- Complete the script *hw9_pb4.py* to plot the projection of the data onto the two first principal components and the two last ones. Submit your plots and comment on the differences.
 - Imagine you have a training set and a test set. Suggest a way to classify the seeds of the test set just using dot products and the svd.

¹The data can be found at <https://archive.ics.uci.edu/ml/datasets/seeds>