DSC 190 - Homework 04
Due: Wednesday, April 27

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 11:59 PM.

Problem 1. (Extra Credit)

We'd like to hear how things are going. Please consider filling out the feedback form at

$$\texttt{https://forms.gle/i5UyG9SA3oBWmJnN8}.$$

Your feedback is especially important for this class, since it is only the second offering. Your honest opinion will help make this class better. The feedback form is totally anonymous (as long as you don't enter your email!)

If you fill out this form, respond to this question and let us know and you'll get 2 points of extra credit towards this homework.

Problem 2.

As a data scientist, there will be many times when you will be working with massive, high dimensional data sets consisting of hundreds of thousands or even millions of points. This is not one of those times.

In this problem, we'll work with the following data set of three points:

$$x^{(1)} = (1,3)^T$$
$$x^{(2)} = (-3,-9)^T$$
$$x^{(3)} = (2,6)^T$$

a) Compute the sample covariance[1] matrix by hand. Show the calculations for each entry of the matrix.

b) What is the top eigenvector of the covariance matrix? You do not need to calculate the eigenvector explicitly, but you should justify your answer.

Hint: plot the data.

c) What is the eigenvalue associated with the top eigenvector?

d) Reduce the dimensionality of each point above by carrying out PCA by hand. Be sure to use the normalized eigenvector. Show your calculations.

Hint: one of your new features should be equal to $-3\sqrt{10}$.

e) The result of PCA is a data set consisting of three numbers. Compute the variance of these three numbers.

Hint: the result should be familiar.

---

[1] Use the version of the sample covariance defined in lecture, not the one that divides by $n-1$.