CS/Math 350
Spr 2025

Exam 1: Take Home

Instructions:

- This is the take-home portion of Exam 1, which consists of 3 problems, each worth 10 points (the in-class portion was worth 70 points).
- For this part of the exam, **you may work with one other person** (please note who it was on your materials), and each of you will upload solutions to Canvas. You may use the text or anything on our class website. You may also use Octave online, but you may not use anything else on the internet.
- You may use either Matlab (Octave) or Python to complete the questions.
- **What to upload?** For each question, upload the code you used, and the output from the computer (either as screen shots or in a document). Be sure to include comments that finish answering each question (if applicable).

Due date: Saturday, 11:59 PM

1. Some basic statistics and projections: Write a script file that will do the following:
   (a) Create a matrix of data where we have 500 points in $\mathbb{R}^{10}$. The data should have a non-zero mean, and should have positive and negative values, and should be random (`randn` works well for this, but you'll need to translate the data).
   (b) Mean subtract the data.
   (c) Project the mean-subtracted data to the plane in $\mathbb{R}^{10}$ spanned by the vectors (written as rows below):
   $$\boldsymbol{x} = (1,0,1,0,1,0,1,0,1,0), \quad \boldsymbol{y} = (1,2,3,1,2,3,1,2,3,1)$$
   (d) Plot the two-dimensional (coordinate) data in $\mathbb{R}^2$.
2. Linear regression
   (a) Create a data set with some noise:

   (Matlab Code)
   ```
   x=7*randn(100,1)-2;
   y=3+x+x.^2+0.2*randn(100,1);
   ```

   (Python Code)
   ```
   import numpy as np
   x = 7 * np.random.randn(100, 1) - 2
   y = 3 + x + x**2 + 0.2 * np.random.randn(100, 1)
   ```

   (b) With the model equation being: $\boldsymbol{y} = \beta_0 + \beta_1 x + \beta_2 x^2$, construct the design matrix A for the linear regression problem.
   (c) Solve the equation $A\boldsymbol{c} = \boldsymbol{y}$ using the normal equations.
   (d) Construct the reduced SVD of A.
   (e) From the SVD, construct the pseudoinverse of A, and then solve the equation $A\boldsymbol{c} = \boldsymbol{y}$.
   (f) Be sure and print your values for $\beta_0, \beta_1, \beta_2$.
3. A Best Basis for the Space of Faces.

   Before we state the problem, remember that we had a sample Matlab file and a sample Python file given in the "Eigenfaces" discussion page from our class website. You should download this code to get you started, then you just need to make a few modifications.

In this question, we'll use data that represents 30 photographs of undergraduate students from Whitman. Each photograph is an array of 294×262 pixels each, and they are stored as vectors in $\mathbb{R}^{77028}$. Therefore, when you type `load Faces` you will have a matrix Y that is 77028×30. Below we'll change the data type of Y (if you know what that is- it's fine if you don't) to a "float" so that we can compute things like the mean.

(Matlab code)

```
load Faces.mat
Y=double(Y);
m=294; n=262;
```

(Python code)

```
import numpy as np
import scipy.io

mat_contents = scipy.io.loadmat('Faces.mat')
Y=mat_contents['Y'];
Y=np.array(Y,dtype=np.float64)
m=294
n=262
boys=mat_contents['boys']
girls=mat_contents['girls']
```

(The Faces.mat file is on our class website).

You will also see two other vectors, `boys` and `girls`. The vectors contain the indices for the photos of boys and girls, respectively (so that `Y(:,girls)` would contain the data only for the girls, for example). We won't be using these vectors here.

We want to find the best basis for the space of faces in this database- That is, the best basis in $\mathbb{R}^{77028}$- Just like we did for the sequence of movie frames. In particular, you should:

(a) Find the mean face and visualize it in Figure 1.

(b) Mean-subtract your data (your mean is a vector in $\mathbb{R}^{77028}$), and construct the economy sized SVD.

(c) Find the best basis vectors via the SVD. Visualize the first four "eigenfaces" in Figure 2.

(d) Choose a face at random, and construct the 5, 10 and 15 dimensional reconstructions. Plot the corresponding images in Figure 3 (using `subplot`).

(e) Project your faces to the best two dimensional representation (you should get a matrix that is 30×2 or 2×30) and plot the result in Figure 4.