Modeling Spr 23, Take Home Exam 2

INSTRUCTIONS

- You may work with one other person (both of you should upload solutions). You may use our text, class notes, or anything on our current class page. Do not use anything else on the internet (ask me first if you need something and you're not sure).
- All of the problems here may be solved using Matlab or Octave-online.
- What to turn in? Each problem will have a script file with your commands on it, and output from the computer. You may either publish the script (it would have both script and output), or submit the script and screenshot of your output.
- Please upload your solutions (Matlab scripts and screenshots, or published scripts) to Canvas no later than **Sunday, Apr 16, 11:59PM**. **No late exams will be accepted**.
- Be sure you start early enough so that if you get stuck, or have technology issues, you still have enough time to ask me.

Here are the exam questions:

1. The Best Basis:

   The data we'll be using is in `Problem01Data.mat`. The data consists of 26 photographs, each is 162×149. Therefore, the matrix X that is loaded will be 24138×26.

   (a) Find the mean photo and visualize it as a photo.

   (b) Find the best two dimensional basis for the data, and project the data to that two dimensional space. Plot the results.

   NOTE: You might use the lab from Week 7 as a starting template, since we did almost the identical thing there.

2. Data Clustering 1 (k-means)

   In this example, we'll be using data clustering to "cluster" the number of colors being used in a photograph to 25. In this example, the photograph will be $256 \times 256 \times 3$, so we can think of the data as 65536 points in $\mathbb{R}^3$. The idea is that, if we cluster the data into 25 groups, then the cluster centers (as points in $\mathbb{R}^3$) will represent the color for that group.

   Download the script file `Problem02.m` as a template for this problem. It also has some commands already written for you.

   Once you load the data from `Problem02Data.mat`, then a matrix X will load ($256 \times 256 \times 3$) into memory. First, you'll need to "reshape" that into a matrix (as outlined above), next you'll use k-means clustering on the data (with 25 clusters).

You should call the index output `idx` and the centers being output as `C`, then you won't need to change anything in the "visualization routine".

3. Data Clustering 2 (DBSCAN):

   Use DBSCAN to cluster the data in `Problem03Data.m`. Notice that this data file is an m-file, so you only need to type `Problem03Data` in the command window, and the matrix X will be constructed to be about 4400 points in $\mathbb{R}^2$. For DBSCAN, we'll need `MinPts` and `epsilon`- For data in $\mathbf{R}^2$, it is common to use `MinPts=4`. For `epsilon`, it can be kind of tricky- I'll give you a hint that a good value of `epsilon` will be somewhere between 3 and 5 in this problem. Try experimenting a bit to see if you can get a good value for it.

   I would recommend using the DBSCAN script file `app1DB.m` as a starting template (from the clustering lab from March 27).

4. Gradient Descent:

   Consider the matrix equation given below:

$$A\mathbf{x} = \mathbf{b} \qquad \Rightarrow \qquad \begin{bmatrix} 3 & 1 \\ 2 & 0 \\ 1 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

   This equation does not have an exact solution, so we look for the least squares solution. That is, the solution that minimizes the quantity

$$E(x,y) = \|A\mathbf{x} - \mathbf{b}\|^2$$

   (a) Before we go to the computer, we need an expression for the error, so expand the given quantity in terms of x, y. To help you, I've written the first term:

   $E(x,y) = (3x + y - 2)^2 +$ fill in the rest

   (b) Find expressions for the gradient of E (that is, the partial derivatives E_x and E_y).

   (c) Use the provided function, `Problem04.m` as a template to solve for x, y using Gradient Descent. There are only a few things that you need to provide, they are marked with `???`. Please be sure to read the code to be sure you're being consistent.