Math 401 - Groupwork #5
Curve Fitting and Team Ranking

Instructions: Work through the problems below in order. Write down all of your answers on paper, to be turned in at the end of class. I do not want you to turn in any MATLAB code. Make sure that the name of every student in the group is on the paper. Some problems ask you to save images and email them to me (`ayashins@math.umd.edu`) at the end. Do as many of the problems as you can, and do not stress about finishing all of them. You will receive a grade based on the correctness of what you've turned in, **provided that everyone is actively working on it during class.**

1. In ELMS, download the file `wave.m` to your computer, and then open the script in MATLAB. The script contains 300 (t, y) data points and then makes a scatter plot of them. For what follows, it is recommended that you add your commands directly into the script, and then run the script every time you want to execute them. This will be useful towards the end, when you need to keep making adjustments.

   (a) Run the script to get a look at the data. Since it appears to oscillate like a wave, it is natural to try to fit a curve involving sines and cosines to the data. Find the least-squares curve of the form

   $$y = \beta_1 + \beta_2 \cos t + \beta_3 \sin t$$

   which best fits the data. Write down your function which best fits the data, and also give the associated least-squares error (Note: the error should be bad, well over 50). Uncomment the three lines at the end and produce a graph of your least-squares curve on top of the scatter plot. Save this image as `wave1a.bmp`

   MATLAB tips: Your matrix A for the least-squares problem will be very large. You do not have to (and should not attempt to) enter it manually. Note that you can apply the transpose to turn row vectors into column vectors. You can type commands like `cos(t)` and it will return the row vector whose entries are the cosines of the entries of `t`. If you want to create a vector of ones, you can do so with the `ones` command. For example, `ones(300,1)` returns a 300×1 column vectors whose entries are all 1. Lastly, say `u, v, w` are each 300×1 vectors that we have stored in MATLAB. Then typing `A = [u v w]` will create the 300×3 matrix whose columns are `u,v,w`.

   (b) The problem with the previous model was that the frequency of the sine/cosine used in our model does not match the frequency of the waves that appear in the data (equivalently, the wavelengths do not match.) A better approach would be to find a best fit of the form

   $$y = \beta_1 + \beta_2 \cos(\omega t) + \beta_3 \sin(\omega t).$$

   Briefly explain why there is an issue with attempting to compute $\beta_1, \beta_2, \beta_3, \omega$ to best fit this data using least-squares methods.

   (c) Make an educated guess for an appropriate value of the frequency ω, and then have your script solve for $\beta_1, \beta_2, \beta_3$, plot the graph, and give the least-squares error. Keep adjusting ω and repeating until you are pleased with the result. You should be able to get the least-squares error under 4. Write down the best fit $y = \beta_1 + \beta_2 \cos(\omega t) + \beta_3 \sin(\omega t)$ that you found, the associated least-squares error, and save a picture of the graph as `wave1c.bmp` (In your script, have a line at the top defining the frequency, such as `omega = 1;` Then use `omega` appropriately throughout the script. When you want to tweak the value of `omega`, you can then just do it on the line where `omega` was defined.)

2. Each NFL division has four teams. During the course of a season each pair of teams in a division plays each other twice (and each team gets to be the home team for one of those games.) The NFC South division consists of the Atlanta Falcons (ATL), the Carolina Panthers (CAR), the New Orleans Saints (NO), and the Tampa Bay Buccaneers (TB). The ranking of the division (which is based on total win-loss record, followed by certain tie-breaking rules) at the end of the 2017 season was

Team	Win–Loss (Total)	Win–Loss (Division)
New Orleans (NO)	11–5	4–2
Carolina Panthers (CAR)	11–5	3–3
Atlanta Falcons (ATL)	10–6	4–2
Tampa Bay Buccaneers (TB)	5–11	1–5

New Orleans and Carolina had the same overall record, but New Orleans won the division because they defeated Carolina in both games that they played. Let's apply our team ranking method to rank these four teams based solely on the margins of victory in the twelve division games that were played. The final scores are shown below. Note the @ symbol indicates the home team.

NO	@ CAR
34	13

CAR	@ TB
17	3

TB	@ NO
10	30

ATL	@ CAR
17	20

TB	@ ATL
20	34

CAR	@ NO
21	31

NO	@ ATL
17	20

ATL	@ TB
24	21

ATL	@ NO
13	23

TB	@ CAR
19	22

CAR	@ ATL
10	22

NO	@ TB
24	31

(a) Let's order the teams alphabetically, so that ATL, CAR, NO, TB, have rankings r_1, r_2, r_3, r_4 respectively, and let $\mathbf{r}$ denote the vector whose entries are the rankings. Write down the (inconsistent) linear system $A\mathbf{r} = \mathbf{b}$ that we would ideally like to solve to find the team rankings. Include the "normalization" condition that the sum of the rankings should be 0 (so A should have 13 rows).

(b) Find the least-squares solution and give the numerical values of the rankings. List the teams from best to worst. Does it match the order of the division rankings above?

(c) Suppose Carolina and Atlanta play one more game in the playoffs (after the regular season has concluded.) Who is expected to win, and by how much?

(d) Suppose that there was a 13th division game in which the team with the worst ranking gets to play the team with best ranking one more time. By what amount must the worst team win this game by in order to become the new best ranked team? Assume that scores must be integers.

(e) (Forget the 13th division game from the previous part now.) Let's factor in homefield advantage. We'll make the assumption that the team with homefield advantage is expected to score 4 more points than it would if the game were on a neutral field. For example, the first game listed above would now be interpreted as

$$(r_2 + 4) - r_3 = -21.$$

Carolina has homefield advantage, so we expect them to perform 4 points better than their ranking indicates. Simplifying would give $r_2 - r_3 = -25$ (so it's like they would have lost by 25 if no side had homefield advantage.) Redo the team rankings while taking homefield advantage into consideration.

(f) Suppose Carolina and Atlanta play one more game in the playoffs in Carolina. Using the homefield advantage model, who is expected to win, and by how many points? What if the game were played in Atlanta?