MS&E 114/214 Autumn 2024 Homework 3

Submission is due **12:00 pm Oct 27.** When submitting, please attach the relevant Python scripts and/or Excel files to enable us to give partial credits, and share the link to your Colab and/or Excel files.

Problems 1 and 2

Review the google colab file that we developed in class for regressions. Then solve the two problems in

`https://colab.research.google.com/drive/1i6O6pZYsVyYm1ZulysHOh5mQapQYE_wt#scrollTo=Zhku2Pq528-z`.

Problem 3

(a) You want to solve a linear regression problem while minimizing the maximum error on any of your training points. Formulate this as a Linear Program.

(b) Comment on which of the three error functions we have seen (absolute value error, squared error, and max-error) is the most susceptible and least susceptible to outliers.

(c) Open ended: Give an example where we might have additional constraints on our coefficients in a regression problem.

(d) Open ended: We have used the California housing data set for problems 1 and 2. Read the description of this dataset at `https://scikit-learn.org/stable/datasets/real_world.html#california-housing-dataset`. Write a very brief summary. Provide another neighborhood feature that you think might be useful as a predictor, as well as another interesting target variable that you think you might be able to predict using this data.

Problem 4

Consider a generic social choice problem with n voters and $m < n$ candidates with each voter having a preferred ranking over all the candidates. We are interested in studying different scoring-based aggregation rules to determine the winner in which a score of s_k is given to a candidate ranked at the k^{th} position and the candidate with highest score wins.

Scoring-based rules: A scoring rule is determined by a sequence of non-negative real numbers $(s_1, s_2, \ldots, s_m)$ satisfying $1 = s_1 \geq s_2 \geq \ldots \geq s_m = 0$. A candidate who is ranked at k^{th} position by a voter i is assigned score of s_k by voter i and the winner is decided by calculating the cumulative score for each candidate by summing across all voters. Ties of scores may be broken in an arbitrary way. An example is described below for two different scoring vectors.

Example 1. *Consider an election with three candidates A, B and C and three voters. The first voter ranks $A > C > B$, the second candidate ranks $B > A > C$ and the third candidate ranks $B > A > C$. We now discuss the winner under two different scoring vectors.*

- *Suppose a scoring vector of $(1, 3/4, 0)$ is used. In this case, candidate A gets a cumulative score of $1 + 3/4 + 3/4 = 2.5$, candidate B a cumulative score of $2 + 0 = 2$ and candidate C a cumulative score of $3/4$ and thus A is the winner.*

- *Suppose a scoring vector of* $(1, 1/4, 0)$ *is used. In this case, candidate* A *gets a cumulative score of* $1 + 1/4 + 1/4 = 1.5$*, candidate* B *a cumulative score of* $2 + 0 = 2$ *and candidate* C *a cumulative score of* $1/4$ *and thus* B *is the winner.*

Thus, different scoring vectors may lead to different winners.

We are now ready for the actual problem. Each voter has provided you with a (strict and complete) ranking of the candidates. Your goal is to determine whether there is a scoring-based rule that can make a given target candidate T a winner of the election (either in first place outright, or tied for first place.)

(a) Model this problem as a Linear Program. Clearly describe what the parameters are, what the decision variables are, what the objective is, and what the constraints are. No credit for work that is not clearly legible, or where the LP model is not defined crisply with proper notation.

(b) (EC for 114, required for 214) Solve it for the attached file `vote_ranking.xlsx` for each of the candidates as the target candidate. You can use any solver you like. For your convenience, we are also providing the number of times each candidate is ranked 1, 2, 3, 4, or 5 (in case it is useful to you):

Candidate	**Number of rank 1 votes**	**Number of rank 2 votes**	**Number of rank 3 votes**	**Number of rank 4 votes**	**Number of rank 5 votes**
A	5	6	5	3	7
B	2	2	6	7	9
C	15	7	2	2	0
D	4	3	8	6	5
E	0	8	5	8	5

(c) What would the answer be if $s_1 = 1$ or $s_m = 0$ were not required?

Problem 5 (EC for 114 and 214)

You are given two triangles in the 2-D plane, each specified by the (x,y) coordinates of its three vertices. Write a quadratic program to find the minimum distance between two points which lie in the first and second triangle respectively. For clarity, we say that a point lies in a triangle if the point is in the interior, on an edge, or at a vertex of the triangle.