15.095: Machine Learning Under a Modern Optimization Lens **Fall 2025**

Homework 5: Due November 20

Hand in: **pdf** upload to Gradescope. Please append any Julia code **at the end** of the whole pdf. Please write your collaborators' full names at the top of your homework.

Note: this homework covers Lectures 16-18. We recommend attempting questions after the relevant content has been covered in class.

Question 1: True/False (30 points)

Please classify the following statements as true or false and justify your answer. If the statement is false, please provide a counter example. We will assign 2 points for correctly classifying the answer, and 4 points for the validity of the justification/counterexample.

(a) In the optimal design of experiments setting, suppose that instead of minimizing the max discrepancy of the first and second moments, we minimize the average discrepancy; namely we want to solve the problem min $\sum_{q=2}^{m}\sum_{p=1}^{q-1}|\mu_p(x)-\mu_q(x)|+\rho|\sigma_p^2(x)-\sigma_q^2(x)|$. This can be formulated as a linear optimization problem with $\frac{m(1+2n-m)}{2}$ binary variables and 1 continuous variable.

(b) Optimization should be preferred over randomization independently of the size of the population group.

(c) Suppose that we want to identify exceptional responders to two drugs; namely we have 3 groups (no treatment, treatment 1 and treatment 2) and we want to find a subset of the population that responds well to any of the two treatments. This can be achieved by modifying the formulation we saw in lecture, while still having $\mathcal{O}(n^2)$ variables.

(d) The methodology for identifying Exceptional Responders is particularly useful in clinical trial of drugs that are found to be successful.

(e) The ROAD framework, similarly to other Causal Inference frameworks, makes the assumption that there is no unobserved confounding.

Question 2: The Power of Optimization Over Randomization (30 points)

In this question, let us investigate how an MIO approach compares with other random allocation methods in assigning subjects to different groups to minimize discrepancies. Specifically, we consider a dataset of 20 patients (from `data_diabetes.csv`) and aim to divide them into two groups of equal size. Our objective is to minimize the discrepancy in the mean Body Mass Index (BMI) between the two groups. Please preprocess the data following the steps in Lecture 14, specifically:

$$w_i' = (y_i - \hat{\mu})/\hat{\sigma}, \quad \text{where} \quad \hat{\mu} = \sum_{i=1}^{n} y_i/n \ \text{ and } \ \hat{\sigma}^2 = \sum_{i=1}^{n}(y_i - \hat{\mu})^2/n.$$

- (a) (20 points) Suppose we want to split the numbers into two groups (with 10 numbers in each group) to minimize the discrepancies in centered first and second moments, as we have learned in class. Please implement the following algorithms. Report the total discrepancy ($|\mu_p(x) - \mu_q(x)| + |\sigma_p^2(x) - \sigma_q^2(x)|$) and the discrepancies in the centered first and second moments of each algorithm. Please report results with random seed 15095. **Please discuss the results**.

  1. Randomization: Shuffle all numbers. Split the first half to the first group and the rest to the second group.
  2. Re-randomization: Do randomization 10000 times and choose the one with the lowest sum of the absolute difference of the first and second moments in the two groups.
  3. Pair Matching: Rank all numbers. For every continuous two numbers, randomly split them into the first group or the second group.
  4. Optimization: Use the formulation in lecture notes. Solve this mixed integer optimization problem. Please set $\rho = 0.5$.

- (b) (10 points) Suppose we want to split the data into two groups (with 10 numbers in each group) to minimize the pairwise difference between the numbers in each group. In this case, our objective function is

$$\min \sum_{i \in \text{ group } 1,\ j \in \text{ group } 2} \left|w_i' - w_j'\right| \tag{5.1}$$

  Please formulate (5.1) as a mixed-integer linear optimization problem and implement it in Julia. Report the value of the objective function in (5.1) for this optimization approach and also for the Randomization approach in (a) 1. **Please discuss the results**.

Question 3: R.O.A.D. (40 points)

In this question, we explore and implement the R.O.A.D. methodology in the provided starter notebook hw5_q3_starter_final.ipynb. All questions and point allocations are in the notebook. All solutions should be written in the notebook as well. Please append the entire notebook as a PDF to your homework submission.