16:960:486 Fall 2024 Homework 1

Due: Wednesday, Sept 25th, 17:30

Instructions

For the written components, write or type your answers clearly on a piece of paper, take a picture/scan, and submit it on Gradescope. Please make sure your photo/scan is readable. You may also write your answer on a tablet and submit a PDF file. For the code components, download the relevant R or Python files and follow the instructions to complete the code. Your completed code should run without error and display an output at the end.

1 Problem 1: Matrix Algebra

Part a (written)

We warm up with some matrix multiplications. Let

$$Y = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \end{pmatrix}.$$

First, write down Y^T, that is, the transpose of the first matrix Y. Next, calculate the following matrix multiplication:

$$YX^T = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ y_{31} & y_{32} \end{pmatrix} \begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \end{pmatrix} = ??$$

Part b (written)

Let $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_m)$ be two vectors (vectors are columns by default). Then, the outer product of u and v is an $n \times m$ matrix of the form

$$uv^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} (v_1, v_2, \ldots, v_m) = \begin{pmatrix} u_1v_1 & u_1v_2 & \ldots & u_1v_m \\ u_2v_1 & u_2v_2 & \ldots & u_2v_m \\ \vdots & \vdots & \ddots & \vdots \\ u_nv_1 & u_nv_2 & \ldots & u_nv_m \end{pmatrix},$$

where we view u as an $n \times 1$ matrix and v^T as a $1 \times m$ matrix.

Now, take the Y and X matrices from part 1 and calculate the following sums of outer products of the columns of X and Y:

$$Y_{*1}X_{*1}^T + Y_{*2}X_{*2}^T = ??$$

Compare your answer with part a.

Part c (written)

For a square matrix M with m rows and m columns, we define its trace $\text{tr}(M) = \sum_{i=1}^{m} M_{ii}$, that is, the summation of all the diagonal elements of M.

Prove the following fact: If A is a matrix with m rows and n columns, and B is a matrix with n rows and m columns, then $\text{tr}(AB) = \text{tr}(BA)$.

2 Problem 2

You are playing rock, paper, scissors with a friend. Your friend claims that they have a secret strategy and, in the 7 games that you played together, they won 4 games, tied 2 games, and lost 1 game.

You want to test the null hypothesis that they are playing at random – in which case each game results in a win, tie, or loss with $\frac{1}{3}$ probability.

Part a (code)

You use the following test statistic: you give 1 point for each win, 0 points for each tie, and -1 point for each loss.

Download the file `hw1_p2a.R` or `hw1_p2a.py` and follow the instructions within. Include the completed file as part of your submission.

Part b (code)

You believe that a win is much more impressive than a tie and so you modify the test statistic: you give 2 points for each win, 0 points for each tie, and -1 point for each loss.

Copy your completed script from the previous part into a new file called `hw1_p2b.R` or `hw1_p2b.py` and modify the code to implement the new test statistic. Include the new file in your submission.

3 Problem 3

Consider an extension of A/B test where you compare 3 different video titles. You observe the following:

1. Title A was shown 160 times (160 impressions) and received 60 clicks/views. (observed CTR = 0.375)
2. Title B was shown 120 times and received 31 clicks/views. (observed CTR = 0.26)
3. Title C was shown 150 times and received 41 clicks/views. (observed CTR = 0.27)

You believe title A is the best title but you want to make sure that this result is not due to chance. You test the null hypothesis that all three titles have the same underlying CTR.

You simulate the null hypothesis by randomly permuting all the impressions, which is equivalent to generating a new set of impressions for title A by drawing without replacement, then generating a new set of impressions for title B by drawing without replacement from the remaining impressions, and then giving the rest to title C.

You use the test statistic of (max observed CTR) − (mean observed CTR).

Part a (code)

Download the file `hw1_p3a.R` or `hw1_p3a.py` and follow the instructions within. Include the completed file as part of your submission.

Part b (written)

Now suppose title A has 1600 impressions, title B has 1200 impressions, and title C has 1500 impressions but the clicks/views are still 60, 31, and 41 respectively. Compare the new p-value with that of part a – is it larger, smaller, or about the same? Why is this the case? Give an intuitive justification.

4 Problem 4 (code)

Write a Monte Carlo algorithm in R or Python to calculate $\sum_{k=1}^{10000} k^{1.5}$. You can use some of the codes used in our lecture.