DSC 40A - Midterm 02

March 3, 2020

Name:

PID:

By signing below, you are agreeing that you will behave honestly and fairly during and after this exam. You should not discuss any part of this exam with anyone enrolled in the course who has not yet taken the exam (this includes posting questions about the exam on Piazza!)

Signature:

Name of student to your **left**:

Name of student to your **right**:

(Write "N/A" if a wall/aisle is to your left/right.)

Instructions:

- Write your solutions to the following problems in the boxes provided.
- Scratch paper is provided at the end of the exam.
- No calculators are permitted, but a cheat sheet is.
- Write your name or PID at the top of each sheet in the space provided.

Tips:

- Show your work to receive partial credit.
- Look through the entire exam before starting.
- Make good use of all assumptions given to you.

(Please do not open your exam until instructed to do so.)

Problem 1.

In each part, select the result of the matrix expression. Hint: in every case, the correct result can be calculated directly, but in most cases it is possible to infer the correct answer using very little calculation.

a) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{pmatrix} =$

○ $\begin{pmatrix} 14 & 32 & 50 \\ 16 & 77 & 122 \\ 25 & 61 & 194 \end{pmatrix}$ ○ $\begin{pmatrix} 14 & 32 & 50 \\ 32 & 77 & 122 \\ 50 & 122 & 194 \end{pmatrix}$ ○ $\begin{pmatrix} 66 & 78 & 90 \\ 78 & 93 & 108 \\ 90 & 108 & 126 \end{pmatrix}$

b) $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & 3 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} =$

○ $\begin{pmatrix} 4 \\ 2 \\ 4 \end{pmatrix}$ ○ $\begin{pmatrix} 4 & 0 & 4 \\ 0 & 2 & 0 \\ 4 & 0 & 4 \end{pmatrix}$ ○ $\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ ○ $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$

c) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 3 \\ 4 \end{pmatrix} =$

○ $\begin{pmatrix} 154 \\ 352 \end{pmatrix}$ ○ $\begin{pmatrix} 154 & 352 \end{pmatrix}$ ○ $\begin{pmatrix} 154 & 352 \\ 352 & 154 \end{pmatrix}$ ○ $\begin{pmatrix} 1 \\ 4 \end{pmatrix}$

d) $\frac{d}{d\vec{x}} \|A\vec{x} - \vec{y}\|^2 =$

○ $2\|A\vec{x} - \vec{y}\|$ ○ $A^T A\vec{x} - A^T\vec{y}$ ○ $2A^T A\vec{x} - 2A^T\vec{y}$

e) $\frac{d}{d\vec{x}} (\vec{x} \cdot \vec{x})^3 =$

○ $3(\vec{x} \cdot \vec{x})^2$ ○ $6\,(\vec{x} \cdot \vec{x})^2\,\vec{x}$ ○ $3\,(\vec{x} \cdot \vec{x})^2\,\vec{x}$ ○ $3\|\vec{x}\|^2$

PID or Name: ______________________________

Problem 2.

Calculate $\frac{d}{d\vec{x}}\|\vec{x}-\vec{y}\|^2$, where $\vec{x}$ and $\vec{y}$ are d-vectors. You may use whatever method you wish, but you should show your work.

Problem 3.

The table below shows data on five universities. In the column labeled "Private", a 1 is used to denote that a university is Private, while 0 is used to denote that it is Public. The tuition shown is the "in-state tuition," where applicable. Tuition and the number of students are listed in multiples of one thousand.

Institution	Private	# of Students	Acceptance Rate	Tuition
Harvard	1	7.5	0.05	50
UCSD	0	35	0.34	14
Carnegie Mellon	1	6.5	0.22	55
UT Austin	0	51	0.36	10.5
University of Virginia	0	22	0.27	18

Suppose we wish to predict a university's tuition by fitting a linear prediction function of the form:

$$\text{Tuition} = w_0 + w_1 \times (\text{Private}) + w_2 \times (\#\text{ of Students}) + w_3 \times (\text{Acceptance Rate})$$

a) Write down, but do not solve, the normal equations for this problem and data set. You do not need to simplify your answer. Keep the number of students and tuition as shown in the table (that is, the number of students for Harvard is 7.5 students).

b) Which of the below is the least squares solution for $\vec{w} = (w_0, w_1, w_2, w_3)^T$?

○ $\begin{pmatrix} 30.34 \\ -32.9 \\ -0.2 \\ 32.8 \end{pmatrix}$ ○ $\begin{pmatrix} -30.34 \\ -32.9 \\ -0.2 \\ 32.8 \end{pmatrix}$ ○ $\begin{pmatrix} -30.34 \\ 32.9 \\ -0.2 \\ -32.8 \end{pmatrix}$ ○ $\begin{pmatrix} 30.34 \\ 32.9 \\ -0.2 \\ -32.8 \end{pmatrix}$

c) Use your answer for the previous part to predict the tuition of the University of Chicago. UChicago is a private school with 14 thousand students and has an acceptance rate of 0.1. You do not need to calculate your answer exactly, but it should be within a few percentage of the exactly-calculated prediction.

d) Suppose the mean squared error of the prediction rule above is E_1. Now suppose that another ten colleges are added to the data set, and a linear prediction rule is fit to the new data; let E_2 be the MSE of this new rule. Is it possible for E_2 to be smaller than E_1?

◯ Yes ◯ No

e) Your friend wishes to make their prediction rule more accurate by using more features, such as the university's average SAT score and the graduation rate. After adding a few features, they notice that the mean squared error of their new prediction has decreased, as expected. They decide to keep adding features until the mean squared error starts to increase again.

Give two different reasons for why this is a bad strategy.

Problem 4.

A hiker has gone missing in the desert, and volunteers are being dispatched to search for them. In this problem, assume that the desert has been divided into 100 distinct search areas. You may leave your answers unsimplified, but they should not contain $\sum$ or $\ldots$.

a) First, it must be decided which of the 100 search areas should actually be searched. How many different ways are there of choosing which areas are searched?

b) Suppose there are 30 volunteers available to search. Each searcher's area will be randomly chosen by generating a number between 1 and 100, with replacement. How many different assignments of searchers to search areas are possible?

c) You have been given a group of six search areas to explore. In how many different orders can these six areas be searched?

d) Suppose now that each of the searchers will be assigned a group of three search areas. How many ways are there to assign groups of areas to searchers? You may assume than an area may have multiple searchers, and the order of the areas in a group does not matter.

e) Suppose now that there are 150 volunteers and 100 search areas. One volunteer will be assigned to each search area by drawing 100 names from a hat, one at a time, without replacement. The remaining 50 people will be assigned to another job and will not be assigned to search areas. How many different assignments of searchers to search areas are possible?

f) Assume that 10 of the 30 volunteers are experienced searchers, while the other 20 are inexperienced. How many ways are there of forming a group of seven searchers in which exactly two of the searchers are experienced?

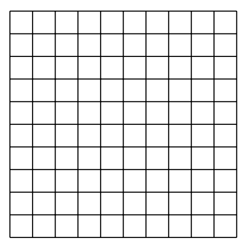
g) The 100 search areas are equally-divided into four regions: North, South, East, and West. How many ways are there to assign groups of three search areas to each of 30 searchers such that all of a searcher's groups are in the same region? Each area may have zero or more searchers assigned to it, and the order of the areas in a group does not matter.

Problem 5.

As in the previous problem, assume that a hiker has gone missing in the desert, and that the desert has been divided into 100 search areas. You may assume that the hiker is in one of the search areas. You may leave your answers unsimplified, but they should not contain $\sum$ or $\ldots$.

a) It is known that 80% of searches last 1 day, 10% of searches last 2 days, 5% of searches last three days, 2% of searches last four days, and 3% of searches last five or more days. What is the probability that a search will last three or fewer days?

b) Suppose that the 100 search areas are arranged in a 10×10 grid, like below:



If you are assigned to a search area at random, what is the probability that you will be assigned to one of the areas on the border of the grid?

c) Suppose that areas are searched one-at-a-time, and that the order in which they are searched is randomly chosen (each order having equal probability). What is the probability that the missing hiker is in the first area searched?

d) You and two of your friends have volunteered to search. If the area each searcher is assigned to is chosen randomly by generating a number between 1 and 100, with replacement, what is the probability that you and your two friends are all assigned to the same area?

e) Suppose that each of 30 searchers will be assigned to a group of 5 by drawing names from a hat. What is the probability that you and your two friends are placed into the same group?

f) Some hikers carry a device called a Personal Locator Beacon (PLB) which allows them to send a signal to rescuers in the event of an emergency. Carrying a PLB is believed to greatly increase your chance of survival.

Suppose 20% of hikers who are found are carrying a PLB, and 1% of hikers who are not found carried a PLB. Assume that 90% of lost hikers are found. What is the probability that a hiker is found given that they carry a PLB?

g) Suppose that 10% of hikers carry a GPS device and 40% carry a compass. Suppose, too, that of those who carry a GPS device, 20% carry a compass, and that of those who carry a compass, 5% carry a GPS device. What is the probability that a randomly-chosen hiker carries at least one of a compass or a GPS?

Before turning in your exam, please check that your name is on every page.

(This is scratch paper. Please remove it from the exam before turning it in.)

(This is scratch paper. Please remove it from the exam before turning it in.)