DSC 40A - Midterm 02

March 3, 2020

Name: SOLUTIONS

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

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

Solution: Instead of performing the total matrix multiplication, we can compute single entries and use them to eliminate some choices. First, entry $(1, 2)$ of the result is the dot product of the first row of the first matrix with the second column of the second:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = 4 + 10 + 18 = 32$$

This means that the last option is incorrect. To differentiate between the first two options, we compute entry $(3, 2)$ of the result by taking the dot product of the third row of the first matrix with the second column of the second matrix. We find that it is 122, and so the right answer must be the second one.

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

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

Solution: The "trick" here is to use the fact that multiplying a matrix by a vector results in a linear combination of the matrix's columns. So, for instance, the result of multiplying any matrix A by the vector $(1, 0, 0)^T$ is the first column of A. So, grouping the rightmost matrix with the vector and performing the multiplication:

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

Repeating this:

$$
\begin{aligned}
&= \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\
&= 1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 3 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 4 \\ 2 \\ 4 \end{pmatrix}
\end{aligned}
$$

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

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

Solution: We can narrow down the options by finding the size of the result. The product of the first two matrices results in a 2×1 matrix (or a column vector). The product of the last two vectors is a scalar. Hence the result is a 2×1 matrix (or column vector). This rules out all but the first and last options. It is clear that the first entry will be much larger than 1, so it cannot be the last option.

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

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

Solution: This is the gradient of the mean squared error; we saw the answer in lecture.

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

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

Solution: Using the chain rule:

$$\frac{d}{d\vec{x}} \left(\vec{x} \cdot \vec{x}\right)^3 = 3 \left(\vec{x} \cdot \vec{x}\right)^2 \frac{d}{d\vec{x}} \left(\vec{x} \cdot \vec{x}\right)$$

We recall that the gradient of $\vec{x} \cdot \vec{x}$ is just $\vec{2}x$. So:

$$\begin{aligned} &= 3 \left(\vec{x} \cdot \vec{x}\right)^2 2\vec{x} \\ &= 6 \left(\vec{x} \cdot \vec{x}\right)^2 \vec{x} \end{aligned}$$

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.

Solution: First, we need to "expand" the norm. Recall that $\|\vec{u}\|^2 = \vec{u} \cdot \vec{u}$, so:

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

Using the distributive property of the dot product:

$$= \frac{d}{d\vec{x}}\left(\vec{x} \cdot \vec{x} - 2\vec{x} \cdot \vec{y} + \vec{y} \cdot \vec{y}\right)$$

Using the linearity of the gradient:

$$= \frac{d}{d\vec{x}}\vec{x} \cdot \vec{x} - \frac{d}{d\vec{x}}2\vec{x} \cdot \vec{y} + \frac{d}{d\vec{x}}\vec{y} \cdot \vec{y}$$

The last term is zero, since it does not involve $\vec{x}$. We remember that the first term is $\vec{x}$. We might remember that the second term is $2\vec{y}$, but we could also compute this by expanding and taking partial derivatives. This leaves us with:

$$= 2\vec{x} - 2\vec{y}$$

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).

Solution: The design matrix is:

$$X = \begin{pmatrix} 1 & 1 & 7.5 & 0.05 \\ 1 & 0 & 35 & 0.34 \\ 1 & 1 & 6.5 & 0.22 \\ 1 & 0 & 51 & 0.36 \\ 1 & 0 & 22 & 0.27 \end{pmatrix}$$

and the observation vector is:

$$\vec{y} = \begin{pmatrix} 50 \\ 14 \\ 55 \\ 10.5 \\ 18 \end{pmatrix}.$$

The parameter vector (which we are solving for) is:

$$\vec{w} = \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

And the normal equations are $X^T X\vec{w} = X^T\vec{y}$.

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

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

Solution: The preferred way to determine the right answer is to interpret the sign of the various elements of $\vec{w}$. Looking at the table, we see that private schools tend to cost more than public schools. As a result, w_1 should be positive. This rules out the first two options.

Therefore we need to choose between the last two options. We can do this by "testing" the ability of each to make good predictions. For instance, the second-to-last option predicts that Harvard's tuition is:

$$(1)(-30.34) + (32.9)(1) + (-.2)(7.5) + (-32.8)(.05) \approx 0$$

on the other hand, the last option predicts:

$$(1)(30.34) + (32.9)(1) + (-.2)(7.5) + (-32.8)(.05) \approx 60$$

So we feel that the last option is the best.

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.

Solution: We take the dot product of $\vec{w}$ with the feature vector for UChicago:

$$\begin{pmatrix} 30.34 \\ 32.9 \\ -0.2 \\ 32.8 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 14 \\ 0.1 \end{pmatrix} = (1)(30.34) + (1)(32.9) + (-.2)(14) + (.1)(32.8) \approx 62$$

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

Solution: This is a repeat of question 3b on the first midterm; please see the explanation there.

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.

Solution: It is a bad strategy, because adding a feature will never cause the MSE to increase. Your friend will continue adding features until they run out of features to add (or the MSE becomes zero).

It is also a bad strategy because it is likely to overfit.

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?

> **Solution:** 2^{100}
>
> There are two choices for each area: either search it, or don't. Thus 2^{100}.

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?

> **Solution:** 100^{30}

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

> **Solution:** $6!$

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.

> **Solution:** $\binom{100}{3}^{30}$

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?

> **Solution:** $\frac{150!}{50!}$

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?

Solution: $\binom{10}{2}\binom{20}{5}$

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.

Solution: $\left[4 \cdot \binom{25}{3}\right]^{30}$

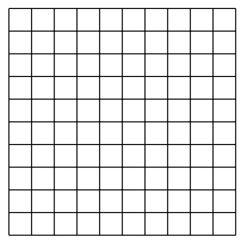
Problem 5.

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?

Solution: 0.95

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?

Solution: There are $100 - (8 \times 8) = 100 - 64 = 36$ border areas. Therefore the probability is 0.36.

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?

Solution: 1/100

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?

Solution: $\frac{100}{100^3} = \frac{1}{100^2}$

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?

Solution: $\frac{\binom{27}{2}}{\binom{29}{4}}$

For the sample space, we'll use the set of all groups which contain you. There are $\binom{29}{4}$ ways to choose the other four members of your group, and all are equally likely.

How many ways are there to form a group that contains you and your two friends? There are two open slots, and to fill them we choose from 27 people. Hence there are $\binom{27}{2}$ such groups.

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?

Solution: We have $P(\text{PLB} \mid \text{Found}) = 0.2$, $P(\text{PLB} \mid \text{Not Found}) = 0.01$, and $P(\text{Found}) = 0.9$. We want $P(\text{Found} \mid \text{PLB})$. Let's use Bayes' Theorem:

$$P(\text{Found} \mid \text{PLB}) = \frac{P(\text{PLB} \mid \text{Found}) \cdot P(\text{Found})}{P(\text{PLB})}$$

Using the law of total probability:

$$= \frac{P(\text{PLB} \mid \text{Found}) \cdot P(\text{Found})}{P(\text{PLB} \mid \text{Found})P(\text{Found}) + P(\text{PLB} \mid \text{Not Found})P(\text{Not Found})}$$

We can say that $P(\text{Not Found}) = 1 - P(\text{Found}) = 0.1$. Therefore we can fill in all of the probabilities with numbers:

$$= \frac{(0.2)(0.9)}{(0.2)(0.9) + (0.01)(0.1)}$$

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?

Solution: P(GPS) = .1, P(compass) = .4, P(compass | GPS) = .2

$$\begin{aligned} P(\text{compass} \cup \text{GPS}) &= P(\text{compass}) + P(\text{GPS}) - P(\text{compass} \cap \text{GPS}) \\ &= P(\text{compass}) + P(\text{GPS}) - P(\text{compass} \mid \text{GPS})P(\text{GPS}) \\ &= .4 + .1 - (.2)(.1) \\ &= .48 \end{aligned}$$

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