Assignment #0

Due: never

Welcome to CS181!

The purpose of this assignment is to assess your readiness for this course. **It will not be graded**. Nonetheless, we *strongly* encourage you to look through this document, and compare your solutions to the provided answer key. Note: this assignment is substantially longer than the homeworks you will receive in the class. Do not feel the need to complete every last problem. Rather, use this as an opportunity to identify areas where you may be a little rusty, and take time now to address any gaps.

If you encounter any difficulty with these problems, fear not! Swing by office hours, post questions on Ed, and consult the resources recommended in this document. We recognize that the pre-requisites for this coarse are broad, and not every student will be freshly familiar with every single topic. That being said, there will be little time — if any — devoted to covering these topics during lecture. If you struggle with the majority of these questions, please consider taking this course at a later semester.

1 Linear algebra (and some calculus)

Linear algebra is foundational for many of the topics we will study this semester. Feeling rusty? Took linear algebra one-too-many semesters ago? Not a problem! Here are some resources to get you back up to speed:

- The Essence of Linear Algebra: an excellent tour of linear algebra basics, brought to you by 3Blue1Brown.
- Interactive Linear Algebra: a succinct, clear text on linear algebra fundamentals, with embedded interactive demos that help illustrate important concepts.
- Matrix Cookbook: the indispensable reference for all things matrix-related. Of particular use are the matrix derivative formulas (chapter 2).

The following questions are representative of the level of linear algebra expected in this course.

Problem 1

Given the matrix $\mathbf{X}$ and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

(a) Expand $\mathbf{X}\mathbf{y} + \mathbf{z}$.

(b) Expand $\mathbf{y}^{\mathbf{T}}\mathbf{X}\mathbf{y}$.

Problem 2

Assume matrix $\mathbf{X}$ has shape $(n \times d)$, and vector $\mathbf{w}$ has shape $(d \times 1)$.

(a) What shape is $\mathbf{y} = \mathbf{X}\mathbf{w}$?

(b) What shape is $(\mathbf{X}^T\mathbf{X})^{-1}$?

(c) Using $\mathbf{y}$ from part (a), what shape is $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$?

(d) Assume vector $\mathbf{w}' = \mathbf{w}^T$. What shape is $\mathbf{y}' = \mathbf{X}\mathbf{w}'^T$?

Problem 3

Write $\mathbf{u} = \mathbf{u}^{\parallel} + \mathbf{u}^{\perp}$ where $\mathbf{u}^{\parallel} = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$. is the projection of $\mathbf{u}$ onto $\mathbf{v}$. Prove that $\langle \mathbf{u}^{\parallel}, \mathbf{u}^{\perp} \rangle = 0$ and that $\mathbf{u} = \mathbf{u}^{\parallel}$ if and only if $\mathbf{u}$ is a scaled multiple of $\mathbf{v}$.

Problem 4

For an invertible matrix $\mathbf{A}$ show that $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$ where $|\mathbf{A}|$ is the determinant of $\mathbf{A}$.

Problem 5

Solve the following vector/matrix calculus problems. In all of the below, $\mathbf{x}$ and $\mathbf{w}$ are column vectors (i.e. $n \times 1$ vectors). It may be helpful to refer to *The Matrix Cookbook* by Petersen and Pedersen, specifically sections 2.4, 2.6, and 2.7.

(a) Let $f(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$. Find $\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\delta}{\delta \mathbf{x}} f(\mathbf{x})$.

Hint: As a first step, you can expand $\mathbf{x}^T \mathbf{x} = (x_1^2 + x_2^2 + ... + x_n^2)$, where $\mathbf{x} = (x_1, ..., x_n)$.

(b) Let $f(\mathbf{w}, \mathbf{x}) = (1 - \mathbf{w}^T \mathbf{x})^2$. Find $\nabla_{\mathbf{w}} f(\mathbf{w}, \mathbf{x}) = \frac{\delta}{\delta \mathbf{w}} f(\mathbf{w}, \mathbf{x})$.

(c) Let $\mathbf{A}$ be a symmetric n-by-n matrix. If $f(\mathbf{w}, \mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{w}^T \mathbf{x}$, find $\nabla_{\mathbf{x}} f(\mathbf{w}, \mathbf{x}) = \frac{\delta}{\delta \mathbf{x}} f(\mathbf{w}, \mathbf{x})$.

Problem 6

In her most recent work-from-home shopping spree, Nari decided to buy several house plants. She would like for them to grow as tall as possible, but needs your calculus help to understand how to best take care of them.

(a) After perusing the internet, Nari learns that the height y in mm of her Weeping Fig plant can be directly modeled as a function of the oz of water x she gives it each week:

$$y = -3x^2 + 72x + 70$$

Is this function concave, convex, or neither? Explain why or why not.

(b) Solve analytically for the critical points of this expression (i.e., where the derivative of the function is zero). For each critical point, use the second-derivative test to identify if each point is a max or min point, and use arguments about the global structure (e.g., concavity or convexity) of the function to argue whether this is a local or global optimum.

(c) How many oz per week should Nari water her plant to maximize its height? With this much water how tall will her plant grow?

(d) Nari also has a Money Tree plant. The height y in mm of her Money Tree can be directly modeled as a function of the oz of water x she gives it per week:

$$y = -x^4 + 16x^3 - 93x^2 + 230x - 190$$

Is this function concave, convex, or neither? Explain why or why not.

2 Probability and statistics

This semester's offering of CS181 will place heavier emphasis on probabilistic machine learning. Fluency with basic probability and statistics is essential. Please take extra care to ensure you are comfortable solving the following problems. Some resources that may help:

- A Probability Primer: An accessible, succinct tour of essential ideas in probability and statistics, written for a scientific audience.
- Bayes'Theorem: An entertaining illustration of Bayes' Theorem, brought to you by Veritasium.
- List of probability distributions: Common (and uncommon) probability distributions, helpfully compiled by Wikipedia. Take full advantage of Wikipedia (and Google) when referencing PDFs, CDFs, moments, and hard-to-remember properties.

The following questions are representative of the level of probability and statistics expected in this course.

Problem 7
Solve the following:

(a) Verify that $\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$.

(b) Verify that $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$.

(c) Verify that $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

(d) Verify that $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Cov}(X, Y)$

(e) Suppose that $X_1, ..., X_n$ are i.i.d., scalar random variables with mean μ and variance σ^2. Let $\bar{X}$ be the mean $\frac{1}{n}\sum_i^n X_i$. Find $\mathbb{E}(\bar{X})$ and $\mathrm{Var}(\bar{X})$.

Problem 8
Suppose $X_1, X_2, X_3, \ldots X_n \overset{\text{iid}}{\sim} \mathrm{Unif}[0, 1]$. What is the distribution of

(a) (X_1, X_2)

(b) X_1X_2

(c) $X_1 + X_2$

(d) $X_1 + X_2 + X_3$

(e) $\sum_{i=1}^n X_i$

(f) $\sum_{i=1}^n X_n$

If writing a precise expression proves cumbersome, feel free to offer a qualitative description or plot instead. What do you notice about the sum in part (e) as $n \to \infty$? How does it differ from the sum in part (f)?

Problem 9
Prove or come up with counterexamples for the following statements:

(a) Random variables A and B are conditionally independent given C. Does this imply that A and B are unconditionally independent?

(b) Random variables A and B are independent. Does this imply that A and B are conditionally independent given some random variable C?

Problem 10

Consider the following:

(a) Your child has been randomly selected for Type I diabetes screening, using a highly accurate new test that boasts of a false positive rate of 1% and a false negative rate of 0%. The prevalence of of Type I diabetes in children is approximately 0.228%. Should your child test positive, what is the probability that they has Type I diabetes?

(b) Should you be concerned enough to ask for further testing or treatment for your child?

(c) Later, you read online that Type I diabetes is 6 times more prevalent in prematurely born children. If this statistic is true, what is the probability that your child, who is prematurely born, has Type I diabetes?

(d) Given the new information, should you be concerned enough to ask for further testing or treatment for your child?

Problem 11

During shopping week, you're trying to decide between two classes based on the criteria that the class must have a lenient grading system. You hear from your friends that one of these classes is rumored to award grades lower than the work merits 35% of the time while the other awards lower grades 15% of the time. However, the rumor doesn't specify which class has harsher grading. So, you decide to conduct an experiment: submit an assignment to be graded.

Fortunately, both classes offer an optional Homework 0 that is graded as extra credit. Unfortunately, you only have time to complete the problem set for just one of these classes.

Suppose you randomly pick the Homework 0 from Class A to complete and suppose that you received a grade that you believe is lower than the quality of your work warrents. Based on this evidence, what is the probability that Class A has the harsher grading system? Which class should you drop based on the results of your experiment (or do you not have sufficient evidence to decide)?

Problem 12

Suppose we randomly sample a Harvard College student from the undergraduate population. Let X be the indicator of the sampled individual concentrating in computer science, and let Y be the indicator that they work in the tech industry after graduation.

Suppose that the below table represented the joint probability mass function (PMF) of X and Y:

	$Y = 1$	$Y = 0$
$X = 1$	$\frac{10}{100}$	$\frac{5}{100}$
$X = 0$	$\frac{15}{100}$	$\frac{70}{100}$

(a) Calculate the marginal probability $P(Y = 1)$. In the context of this problem, what does this probability represent?

(b) Calculate the conditional probability $P(Y = 1|X = 1)$. In the context of this problem, what does this probability represent?

(c) Are X and Y independent? Why or why not? What is the interpretation of this?

Problem 13

(Optional) A random point (X, Y, Z) is chosen uniformly in the ball

$$B = \{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$$

(a) Find the joint PDF of (X, Y, Z).

(b) Find the joint PDF of (X, Y) (this is the marginal distribution on X and Y).

(c) Write an expression for the marginal PDF of X, as an integral.

Credits: Problems 12 and 13 were inspired by Exercise 7.19 and Example 7.1.5 in Blitzstein & Hwang's "Introduction to Probability".

3 Programming Exercises

This course uses Python as the primary programming language, and will involve significant coding. If you have never used Python before, it may be possible to learn the language on-the-fly, but doing so will require considerable extra self-study. Some resources for refreshing your Python skills:

- The Python Tutorial: straight from the docs. Chapters 2 - 5 are the most important.
- Automate the Boring Stuff with Python: an excellent overview of the language, as well as useful Python skills everyone should know. Chapters 1 - 5 will teach you everything you need to know. Chapters ≥ 12 are nuggets of pure gold.
- NumPy Quickstart: a succinct introduction to `numpy`, the indispensable numeric Python toolkit. If you've never used `numpy` before, please give this a look and try the practice exercises below.

The following exercises illustrate common programming challenges you will encounter in this class:

Problem 14

Let's warm up with a simple Python exercise:

(a) Using the random seed 181, generate $N = 20$ data points (x, y) where x is uniformly sampled from $(-10, 10)$ and y is uniformly sampled from $(20, 80)$. (To get consistent answers with the staff solution, use `numpy`.)

(b) (Optional) Save your data into two columns in a .csv file, and read it back out.

(c) Let $f(x, y) = (y + 10) \cdot x/5$. Compute $z_i = f(x_i, y_i)$ for $i \in 1 \ldots N$. Report the mean and standard deviation of $\{z_1, \ldots, z_N\}$.

(d) Identify the data point (x, y) with the largest y-value in the data set.

(e) Compute the sum of all the y-values of points with positive x-value.

Problem 15

Here, we will practice using the Python package `numpy`.

(a) Using `np.arange` or `np.linspace`, create a `numpy` array of all the nonnegative integers starting at 0 and ending at 9.

(b) Using `np.reshape`, reshape the array so it's 2 dimensional with size $(2, 5)$. Hint: the first row should be $[0, 1, 2, 3, 4]$.

(c) Using `np.vstack`, add a row to the bottom of the matrix from the previous part so that the matrix contains all of the nonnegative integers starting from 0 and ending at 14 in ascending order left to right, then top to bottom.

(d) Using `np.ones`, `np.reshape`, and `np.hstack`, add a column to the right of the matrix from the previous part that is all ones.

(e) Using `np.dot`, perform a matrix-vector multiplication with the matrix from the previous part with the vector $[0, 1, 0, 0, 0, 0]$.

(f) Using `np.sum` and some of your own logic, find the sum of all the numbers in the matrix from 4. that are even.

Problem 16

In this problem, we'll learn how to use inverse transform sampling to sample from an exponential distribution. The goal of this exercise is to translate a theoretical notion into tangible code — a skill you will practice often in this course.

(a) Let $U \sim \text{Unif}[0, 1]$. Prove that $\Pr(U \leq x) = x$, for any value $x \in [0, 1]$.

(b) Let X be a continuous random variable with a CDF function F that is invertible. Prove that $F(X) \sim \text{Unif}[0, 1]$.

Hint: Suppose there exists an invertible function T such that $T(U) \stackrel{d}{=} X$. Prove that $T(u) = F^{-1}(u)$ for $u \in [0, 1]$.

Hint (x2): If F is the CDF of X, then $F(x) = \Pr(X \leq x)$. Rewrite this expression into a form that looks like the result from part (a).

(c) From part (b), we know that $X \stackrel{d}{=} F^{-1}(U)$. In plain English, this means that if we draw samples $u_1, u_2, u_3, \ldots u_n$ from a uniform distribution, and compute $x_i = F^{-1}(u_i)$, then the resulting values $x_1, x_2, x_3, \ldots x_n$ are sampled according to X!

Why is this a big deal? Why not just call `np.random.exponential()` and save yourself the trouble of going through a laborious derivation? Your computer is very good at sampling random-looking values uniformly between 0 and 1. The challenge in sampling any other distribution is therefore converting from uniform samples to samples of X. If X has an invertible CDF, we can use the inverse transform method to sample from X very cheaply.

Apply your newfound knowledge to sample from the exponential distribution. Write out the CDF of $X \sim \text{Exp}(\lambda)$, compute the inverse, and convert uniform samples into exponential samples. Plot a histogram to confirm that your samples are, indeed, exponentially distributed.