Homework Set #0

Due Sunday 7th April 2024, before 11:59 pm

Submission instructions

- Submit your solutions electronically on the course Gradescope site as PDF files. Refer guide for submission in Handout 1.
- If you plan to typeset your solutions, please use the LaTeX solution template. If you must submit scanned handwritten solutions, please use a black pen on blank white paper and a high-quality scanner.
- We will not grade your submission for homework 0, i.e. your performance on this problem set will not be included in your final grade. However, your homework submission is necessary. **We will not grade your homework 1 and the rest of the homeworks if homework 0 is missing.**

Although many students find a machine learning class to be rewarding, we do assume that you have a basic familiarity with several types of math. Before taking the class, you should evaluate whether you have the mathematical background the class depends upon.

- **Multivariate Calculus** (at the level of a first undergraduate course, *e.g.*, Math 32A and 32B at UCLA). For example, we rely on you being able to take derivatives and integrals. During the class you might be asked, for example, to derive gradients of multivariate functions.
- **Linear Algebra** (at the level of a first undergraduate course, *e.g.*, Math 33A). For example, we assume you know how to multiply vectors and matrices, and that you understand matrix inversion, eigenvectors and eigenvalues. During the class, you might also be asked to also learn about methods for matrix factorization.
- **Probability and Statistics** (at the level of a first undergraduate course, *e.g.*, Statistics 100A, ECE 131A). For example, we assume you know how to find the mean and variance of a set of data, that you are familiar with common probability distributions such as the Gaussian and Uniform distributions, and that you understand basic notions such as conditional probabilities and Bayes rule. During the class, you might be asked to calculate the likelihood (probability) of a data set with respect to some given probability distribution, and to then derive the parameters of the distribution that maximize this likelihood.

This assignment is adapted from course material by Sriram Sankararaman, which is in turn also based on material from William Cohen, Ziv Bar-Joseph (CMU) and Jessica Wu (Harvey Mudd).

This assignment helps you self-evaluate whether you have the background to succeed in the class. For each of these mathematical topics, we provide below (1) a minimum background test and (2) a moderate background test (3) programming skills test. If you pass the moderate background test, you are in excellent shape to take the class. If you pass the minimum background but not the moderate background test, then you can still take the class, but you should expect to devote extra time to fill in necessary math background. If you cannot pass the minimum background test, we suggest you fill in your math background before taking the class.

For programming skills test, you DO NOT have to submit your code. By completing the test you will be able to get a general understanding of how to use python to generate data, plot data and do some basic math operations. If you finish it easily, you are in excellent stage to take this class. If not, please devote some time to familiarize yourself with python and relevant python libraries.

You may find the following resources helpful:

- Andrew Ng's CS229 Course (Stanford)
  - Linear Algebra Review
  - Probability Theory Review
- Numpy/Python
  - Stanford CS 231n Course Python Numpy Tutorial
  - Siddharth Dixit's An Essential Guide to Numpy for Machine Learning in Python

Necessary Minimum Background Test

While you are welcome to use online resources, such as Wolfram-Alpha, you should be able to solve these problems by hand.

Problem 1 (Multivariate Calculus)

Consider $y = x \sin(z) e^{-x}$. What is the partial derivative of y with respect to x?

Problem 2 (Linear Algebra)

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

$$\boldsymbol{X} = \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 1 \\ 3 \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

(a) What is the inner product $\mathbf{y}^T\mathbf{z}$?

(b) What is the product $\boldsymbol{X}\mathbf{y}$?

(c) Is $\boldsymbol{X}$ invertible? If so, give the inverse; if not, explain why not.

(d) What is the rank of $\boldsymbol{X}$?

Problem 3 (Probability and Statistics)

Consider a sample of data S obtained by flipping a coin five times. $X_i, i \in \{1, \ldots, 5\}$ is a random variable that takes a value 0 when the outcome of coin flip i turned up heads, and 1 when it turned up tails. Assume that the outcome of each of the flips does not depend on the outcomes of any of the other flips. The sample obtained $S = (X_1, X_2, X_3, X_4, X_5) = (1, 1, 0, 1, 0)$.

(a) What is the sample mean for this data?

(b) What is the unbiased sample variance ?

(c) What is the probability of observing this data assuming that a coin with an equal probability of heads and tails was used? (*i.e.*, The probability distribution of X_i is $P(X_i = 1) = 0.5$, $P(X_i = 0) = 0.5$.)

(d) Note the probability of this data sample would be greater if the value of the probability of heads $P(X_i = 1)$ was not 0.5 but some other value. What is the value that maximizes the probability of the sample S? [Optional: Can you prove your answer is correct?]

(e) Given the following joint distribution between X and Y, what is $P(X = T|Y = b)$?

$P(X,Y)$		Y		
		a	b	c
X	T	0.2	0.1	0.2
	F	0.05	0.15	0.3

Problem 4 (Discrete and Continuous Distributions)

Match the distribution name to its formula.

(a) Gaussian	(i) $p^x(1-p)^{1-x}$, when $x \in \{0, 1\}; 0$ otherwise
(b) Exponential	(ii) $\frac{1}{b-a}$ when $a \le x \le b; 0$ otherwise
(c) Uniform	(iii) $\binom{n}{x} p^x(1-p)^{n-x}$
(d) Bernoulli	(iv) $\lambda e^{-\lambda x}$ when $x \ge 0; 0$ otherwise
(e) Binomial	(v) $\frac{1}{\sqrt{(2\pi)\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$

Problem 5 (Mean and Variance)

(a) What is the mean and variance of a $Bernoulli(p)$ random variable?

(b) If the variance of a zero-mean random variable X is σ^2, what is the variance of $2X$? What about the variance of $X + 2$?

Problem 6 (Algorithms)

For each pair (f, g) of functions below, list which of the following are true: $f(n) = O(g(n))$, $g(n) = O(f(n))$, or both. Briefly justify your answers.

(a) $f(n) = ln(n), g(n) = lg(n)$. Note that ln denotes log to the base e and lg denotes log to the base 2.

(b) $f(n) = 3^n, g(n) = n^{10}$

(c) $f(n) = 3^n, g(n) = 2^n$

Moderate Background Test

Problem 7 (Probability and Random Variables)

If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]E[Y]$.

Problem 8 (Linear Algebra)

(a) **Vector Norms** [4 pts]
Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with following norms (you can hand draw or use software for this question):

i. $||\mathbf{x}||_2 \leq 1$ (Recall $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$.)
ii. $||\mathbf{x}||_0 \leq 1$ (Recall $||\mathbf{x}||_0 = \sum_{i:x_i \neq 0} 1$.)
iii. $||\mathbf{x}||_1 \leq 1$ (Recall $||\mathbf{x}||_1 = \sum_i |x_i|$.)
iv. $||\mathbf{x}||_\infty \leq 1$ (Recall $||\mathbf{x}||_\infty = \max_i |x_i|$.)

(b) **Matrix Decompositions [6 pts]**

i. Give the definition of the eigenvalues and the eigenvectors of a square matrix.
ii. Find the eigenvalues and eigenvectors of

$$\boldsymbol{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

iii. For any positive integer k, show that the eigenvalues of $\boldsymbol{A}^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$, the k^{th} powers of the eigenvalues of matrix $\boldsymbol{A}$, and that each eigenvector of $\boldsymbol{A}$ is still an eigenvector of $\boldsymbol{A}^k$.

(c) **Vector and Matrix Calculus [5 pts]**
Consider the vectors $\mathbf{x}$ and $\mathbf{a}$ and the symmetric matrix $\boldsymbol{A}$.

i. What is the first derivative of $\mathbf{a}^T\mathbf{x}$ with respect to $\mathbf{x}$?
ii. What is the first derivative of $\mathbf{x}^T\boldsymbol{A}\mathbf{x}$ with respect to $\mathbf{x}$? What is the second derivative?

(d) **Geometry [5 pts]**

i. Show that the vector $\mathbf{w}$ is orthogonal to the line $\mathbf{w}^T\mathbf{x} + b = 0$. (Hint: Consider two points $\mathbf{x}_1, \mathbf{x}_2$ that lie on the line. What is the inner product $\mathbf{w}^T(\mathbf{x}_1 - \mathbf{x}_2)$?)
ii. Argue that the distance from the origin to the line $\mathbf{w}^T\mathbf{x} + b = 0$ is $\frac{b}{||\mathbf{w}||_2}$.

Programming Skills

Start familiarizing yourself with the Python libraries `numpy` and `matplotlib` by completing the following exercises. (You do not have to submit your code.)

You may find the following references helpful:

- https://numpy.org/doc/stable/reference/random/generated/numpy.random.multivariate_normal.html
- http://docs.scipy.org/doc/numpy/reference/generated/numpy.linalg.eig.html

Problem 9 (Sampling from a Distribution)

For questions (a-e), only submit your plots. You do not need to submit code.

(a) Draw 1000 samples $\mathbf{x} = \binom{x_1}{x_2}$ from a 2-dimensional Gaussian distribution with mean $\binom{0}{0}$ and identity covariance matrix, i.e. $p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^2}} \exp\left(-\frac{||\mathbf{x}||^2}{2}\right)$, and make a scatter plot (x_1 vs x_2).

(b) How does the scatter plot change if the mean is $\binom{-1}{1}$?

(c) How does the (original) scatter plot change if you double the variance of each component?

(d) How does the (original) scatter plot change if the covariance matrix is changed to $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$?

(e) How does the (original) scatter plot change if the covariance matrix is changed to $\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$?

Problem 10 (Eigendecomposition)

Write a python program to compute the eigenvector corresponding to the largest eigenvalue of the following matrix and submit the computed eigenvector.

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

.

Problem 11 (Data)

There are now lots of really interesting data sets publicly available to play with. They range in size, quality and the type of features and have resulted in many new machine learning techniques being developed.

Find a public, free, supervised (i.e. it must have features *and* labels), machine learning dataset. You may NOT list a data set from 1) The UCI Machine Learning Repository or 2) from Kaggle.com. Once you have found the data set, provide the following information:

(a) The name of the data set.

(b) Where the data can be obtained.

(c) A brief (i.e. 1-2 sentences) description of the data set including what the features are and what is being predicted.

(d) The number of examples in the data set.

(e) The number of features for each example. If this is not concrete (i.e. it is text), then a short description of the features.

For this question, do not just copy and paste the description from the website or the paper; reference it, but use your own words. Your goal here is to convince the staff that you have taken the time to understand the data set, where it came from, and potential issues involved.