HMC CS 158, Fall 2017
Problem Set 1 Exercises: Math Prerequisites

Goals:
- To help you evaluate whether you have the mathematical background to succeed in the class.
- To explore possible applications of machine learning by finding existing data sets.

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 self-evaluate whether you have the mathematical background the class depends upon.

- **Multivariate Calculus** (at the level of a first undergraduate course). For example, we rely on you being able to take derivatives and integrals.
- **Linear Algebra** (at the level of a first undergraduate course). For example, we assume you know how to multiply vectors and matrices, and that you understand eigenvectors and eigenvalues.
- **Probability and Statistics** (at the level of a first undergraduate course). 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.

For each of these topics, we provide below (1) a minimum background test and (2) a modest background test. If you pass the modest background test, you are in excellent shape to take the class. If you pass the minimum background but not the modest background test, then you can still take the class, but you should expect to devote extra time to fill in necessary math background as the course introduces it. If you cannot pass the minimum background test, we suggest you fill in your math background before taking the class.

You may find the following resources helpful:

- Andrew Ng's CS229 Course (Stanford)
  - Linear Algebra Review (`http://cs229.stanford.edu/section/cs229-linalg.pdf`)
  - Probability Theory Review (`http://cs229.stanford.edu/section/cs229-prob.pdf`)
- Harvey Mudd College Math Department
  - Calculus Tutorials (`https://www.math.hmc.edu/calculus/`)

Additional resources are available on the course syllabus.

Submission

You should submit any answers to the exercises in a single file `writeup.pdf`. This writeup should include your name and the assignment number at the top of the first page, and it should clearly label all problems. Additionally, cite any collaborators and sources of help you received (excluding course staff), and if you are using slip days, please also indicate this at the top of your document.

This assignment is adapted from course material by William Cohen and Ziv Bar-Joseph (CMU).

1 Necessary Minimum Background Test [5 pts]

This should take 10-15 minutes, if you know the material. While you are welcome to use online resources, such as Wolfram-Alpha, you should be able to solve the problems by hand.

1.1 Multivariate Calculus (1 pts)

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

1.2 Linear Algebra (2 pts)

(a) **Matrix Algebra**
Consider the matrix $\boldsymbol{X}$ and the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ below:

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

i. What is the inner product $\boldsymbol{y}^T\boldsymbol{z}$?
ii. What is the product $\boldsymbol{X}\boldsymbol{y}$?
iii. Is $\boldsymbol{X}$ invertible? If so, give the inverse; if not, explain why not.
iv. What is the rank of $\boldsymbol{X}$?

(b) **Vector and Matrix Calculus**
Consider the vectors $\boldsymbol{x}$ and $\boldsymbol{a}$ and the symmetric matrix $\boldsymbol{A}$. (For the following, do not just give the answer – show why.)

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

1.3 Probability and Statistics (2 pts)

(a) Consider a sample of data S obtained by flipping a coin X, where 0 denotes the coin turned up heads, and 1 denotes that it turned up tails: $S = \{1, 1, 0, 1, 0\}$.

i. What is the sample mean for this data?
ii. What is the sample variance ?
iii. 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 is $P(X = 1) = 0.5$, $P(X = 0) = 0.5$.)
iv. Note the probability of this data sample would be greater if the value of $P(X = 1)$ was not 0.5 but some other value. What is the value that maximizes the probability of sample S? [Optional: Can you prove your answer is correct?]

v. 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

(b) **Discrete and Continuous Distributions**
Match the distribution name to its formula.

(a) Multivariate Gaussian	(i) $p^x(1-p)^{1-x}$
(b) Exponential	(ii) $\frac{1}{b-a}$ when $a \leq x \leq b; 0$ otherwise
(c) Uniform	(iii) $\binom{n}{x}p^x(1-p)^{n-x}$
(d) Bernoulli	(iv) $\lambda e^{-\lambda x}$ when $x \geq 0; 0$ otherwise
(e) Binomial	(v) $\frac{1}{\sqrt{(2\pi)^d\|\Sigma\|}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

2 Moderate Background Test [2 pts]

This section of the assignment will be graded for effort. You are not required to attempt every question (especially since some of the questions are very challenging!), but it should be evident that you have used this section to self-evaluate your mathematical fluency.

2.1 Probability and Random Variables (1 pts)

(a) **Probability**
Let A and B be two discrete random variables. In general, are the following true or false? (Here A^c denotes complement of the event A.)

i. $P(A \cup B) = P(A \cap (B \cap A^c))$
ii. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
iii. $P(A) = P(A \cap B) + P(A^c \cap B)$
iv. $P(A|B) = P(B|A)$
v. $P(A_1 \cap A_2 \cap A_3) = P(A_3|(A_2 \cap A_1))P(A_2|A_1)P(A_1)$

(b) **Mean, Variance and Entropy**

i. What is the mean, variance, and entropy of a $Bernoulli(p)$ random variable?
ii. If the variance of a zero-mean random variable X is σ^2, what is the variance of $2X$? What about the variance of $X + 2$?

(c) **Mutual and Conditional Independence**

i. If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]E[Y]$.
ii. Alice rolls a die and calls up Bob and Chad to tell them the outcome A. Due to disturbance in the phones, Bob and Chad think the roll was B and C, respectively. Is B independent of C? Is B independent of C given A?

(d) **Law of Large Numbers and Central Limit Theorem**
Provide one line justifications.

i. If a fair die is rolled 6000 times, the number of times 3 shows up is close to 1000.

ii. If a fair coin is tossed n times and $\bar{X}$ denotes the average number of heads, then the distribution of $\bar{X}$ satisfies
$$\sqrt{n}(\bar{X} - \tfrac{1}{2}) \xrightarrow{n\to\infty} \mathcal{N}(0, \tfrac{1}{4})$$

2.2 Linear Algebra (1 pts)

(a) **Vector Norms**
Draw the regions corresponding to vectors $\boldsymbol{x} \in \mathbb{R}^2$ with following norms:

i. $||\boldsymbol{x}||_2 \leq 1$ (Recall $||\boldsymbol{x}||_2 = \sqrt{\sum_i x_i^2}$.)

ii. $||\boldsymbol{x}||_0 \leq 1$ (Recall $||\boldsymbol{x}||_0 = \sum_{i:x_i \neq 0} 1$.)

iii. $||\boldsymbol{x}||_1 \leq 1$ (Recall $||\boldsymbol{x}||_1 = \sum_i |x_i|$.)

iv. $||\boldsymbol{x}||_\infty \leq 1$ (Recall $||\boldsymbol{x}||_\infty = \max_i |x_i|$.)

(b) **Matrix Decompositions and Rank**

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) **Geometry**

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

ii. Argue that the distance from the origin to the line $\boldsymbol{w}^T\boldsymbol{x} + b = 0$ is $\frac{b}{||\boldsymbol{w}||_2}$. (Hint: Set up an optimization function, and solve using Lagrange multipliers.)

3 Data, Data, Data [1 pts]

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.