HOMEWORK 1
BACKGROUND TEST

CMU 10601: MACHINE LEARNING (SPRING 2016)
OUT: Jan. 13, 2016
DUE: 5:30 pm, Jan. 21, 2016
TAs: William Herlands and Han Zhao

Guidelines

The goal of this homework is for you to determine whether you have the mathematical background needed to take this class, and to do some background work to fill in any areas in which you may be weak. Although most students find the machine learning class to be very rewarding, it does assume that you have a basic familiarity with several types of math: calculus, matrix and vector algebra, and basic probability. You do not need to be an expert in all these areas, but you will need to be conversant in each, and to understand:

- Basic calculus (at the level of a first undergraduate course). For example, we rely on you being able to take derivatives. During the class you might be asked, for example, to calculate derivatives (gradients) of functions with several variables.
- 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 matrix inversion.
- Basic 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, and that you understand basic notions such as conditional probabilities and Bayes rule. During the class, you might be asked to calculate the probability of a data set with respect to a given probability distribution.
- Basic tools concerning analysis and design of algorithms, including the big-O notation for the asymptotic analysis of algorithms.

For each of these mathematical topics, this homework provides (1) a minimum background test, and (2) a medium background test. If you pass the medium background tests, you are in good shape to take the class. If you pass the minimum background, but not the medium background test, then you can still successfully take and pass the class but you should expect to devote some 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. Here are some useful resources for brushing up on, and filling in this background.

Probability

- Lecture notes: http://www.cs.cmu.edu/~aarti/Class/10701/recitation/prob_review.pdf.

Linear Algebra:

- Short video lectures by Prof. Zico Kolter: http://www.cs.cmu.edu/~zkolter/course/linalg/outline.html.
- Handout associated with above video: http://www.cs.cmu.edu/~zkolter/course/linalg/linalg_notes.pdf.
- Book: Gilbert Strang. Linear Algebra and its Applications. HBJ Publishers.

Matlab tutorial

- http://www.math.mtu.edu/~msgocken/intro/intro.pdf.
- http://ubcmatlabguide.github.io/.

Big-O notation:

- http://www.stat.cmu.edu/~cshalizi/uADA/13/lectures/app-b.pdf
- http://www.cs.cmu.edu/~avrim/451f13/recitation/rec0828.pdf
- See ASYMPTOTIC ANALYSIS (Week 1) in the following: https://class.coursera.org/algo-004/lecture/preview

Instructions

- **Submit your homework** on time **BOTH** electronically by submitting to Autolab **AND** by dropping off a hardcopy in the bin outside Gates 8221 by 5:30 pm, Thursday, January 21, 2016.
- **Late homework policy**: Homework is worth full credit if submitted before the due date, half credit during the next 48 hours, and zero credit after that. Additionally, you are permitted to drop 1 homework.
- **Collaboration policy**: For this homework only, you are welcome to collaborate on any of the questions with anybody you like. However, you *must* write up your own final solution, and you must list the names of anybody you collaborated with on this assignment. The point of this homework is not really for us to evaluate you, but instead for *you* to determine whether you have the right background for this class, and to fill in any gaps you may have.

Minimum Background Test [80 pts]

Vectors and Matrices [20 pts]

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

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

1. What is the inner product of the vectors $\mathbf{y}$ and $\mathbf{z}$? (this is also sometimes called the *dot product*, and is sometimes written as $\mathbf{y}^T\mathbf{z}$)
2. What is the product $X\mathbf{y}$?
3. Is X invertible? If so, give the inverse, and if no, explain why not.
4. What is the rank of X?

Calculus [20 pts]

1. If $y = x^4 + 2x^2 - 1$ then what is the derivative of y with respect to x?
2. If $y = \log(\frac{x^7}{10x}) + \sin(z)x^{z-8}$, what is the partial derivative of y with respect to x?

Probability and Statistics [20 pts]

Consider a sample of data $S = \{0, 1, 1, 0, 0\}$ created by flipping a coin x five times, where 0 denotes that the coin turned up heads and 1 denotes that it turned up tails.

1. What is the sample mean for this data?
2. What is the sample variance for this data?
3. What is the probability of observing this data, assuming it was generated by flipping a biased coin with $p(x = 1) = 0.6, p(x = 0) = 0.4$.
4. Note that the probability of this data sample would be greater if the value of $p(x = 1)$ was not 0.6, but instead some other value. What is the value that maximizes the probability of the sample S? Please justify your answer.
5. Consider the following joint probability table where both A and B are binary random variables:

A	B	$P(A, B)$
0	0	0.5
0	1	0.1
1	0	0.3
1	1	0.1

   (a) What is $P(A = 0, B = 0)$?
   (b) What is $P(A = 1)$?
   (c) What is $P(A = 0|B = 1)$?
   (d) What is $P(A = 0 \vee B = 0)$?

Big-O Notation [20 pts]

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.

1. $f(n) = 2^n$, $g(n) = e^n$.
2. $f(n) = n^2$, $g(n) = n^4 + 2n + 3$.
3. $f(n) = n$, $g(n) = \log_{10} n$.

Medium Background Test [20 pts]

Algorithm [5 pts]

Divide and Conquer: Assume that you are given a sorted array with n numbers ranging from $(-\infty, +\infty)$. Additionally, you are told that somewhere in the array is the number 0. Provide an algorithm to locate the 0 which runs in $O(\log(n))$. Explain your algorithm in words, describe why the algorithm is correct, and justify its running time.

Probability and Random Variables [5 pts]

Probability

State true or false. Here Ω denotes the sample space and A^c denotes the complement of the event A.

1. Assume $P(B) > 0$, then $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$.
2. For any $A, B \subseteq \Omega$ such that $P(B) > 0, P(A^c) > 0$, $P(A|B) + P(B|A^c) = 1$.
3. For any $A, B \subseteq \Omega$ such that $0 < P(B) < 1$, $P(A|B) + P(A|B^c) = 1$.
4. For any $A, B \subseteq \Omega$, $P(B^c \cup (A \cup B)) + P(B \cap (A \cup A^c)) = 1$.
5. Let $\{A_i\}_{i=1}^n$ be mutually independent. Then, $P(\bigcap_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

Discrete and Continuous Distributions

Write down the formula of the probabilistic density/mass functions of random variable X.

1. 1d Gaussian distribution. $X \sim \mathcal{N}(x; \mu, \sigma^2)$.
2. Bernoulli distribution. $X \sim \text{Bernoulli}(p), 0 < p < 1$.
3. Uniform distribution. $X \sim \text{Unif}(a, b), a < b$.
4. Exponential distribution. $X \sim \text{Exp}(\lambda), \lambda > 0$.
5. Poisson distribution. $X \sim \text{Poisson}(\lambda), \lambda > 0$.

Mean and Variance

1. Given the following Poisson distribution,

$$p(x; \lambda) = \frac{\frac{\lambda}{3}^x e^{-\lambda/3}}{x!}, \quad \lambda > 0 \tag{1}$$

   (a) What is the mean of the distribution?
   (b) What is the variance of the distribution?

2. Let X be a random variable and $\mathbb{E}[X] = 1, \text{Var}(X) = 1$. Compute the following values:

   (a) $\mathbb{E}[2X]$.
   (b) $\text{Var}(2X)$ and $\text{Var}(X + 1)$.

Mutual and Conditional Independence

1. If X and Y are independent random variables, show that $\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y$. (Hint: $\mathrm{Var}(X+Y) = \mathrm{Var}X + 2\mathrm{Cov}(X,Y) + \mathrm{Var}Y$)
2. Consider rolling two fair, six-sided dice. Let X denote the value of the first die. Let Y denote the value of the second die. Let Z=X+Y.
   (a) Are X and Y independent? That is, if you observe the first die, what do you know about the second die?
   (b) Are X and Y independent given that Z is even?

Law of Large Numbers and the Central Limit Theorem

Provide one line justification.

1. If a fair die is rolled 60000 times, the number of times 1 shows up is close to 10000.
2. Let $X_i \sim \mathcal{N}(0,1)$ and $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the distribution of $\bar{X}$ satisfies

$$\sqrt{n}\bar{X} \stackrel{n\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

Linear Algebra [5 pts]

Vector Norms

Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $||\mathbf{x}||_1 \le 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)
2. $||\mathbf{x}||_2 \le 1$ (Recall that $||\mathbf{x}||_2 = \sum_i \sqrt{\sum_i x_i^2}$)
3. $||\mathbf{x}||_\infty$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$)

Geometry

Prove that these are true or false. Provide all steps!

1. The Euclidean distance from the origin to the hyperplane $\mathbf{w}^T\mathbf{x} + b = 0$ is $\frac{|b|}{||\mathbf{w}||_2}$.
2. The Euclidean distance between two parallel hyperplane $\mathbf{w}^T\mathbf{x} + b_1 = 0$ and $\mathbf{w}^T\mathbf{x} + b_2 = 0$ is $\frac{|b_1 - b_2|}{||\mathbf{w}||_2}$ (Hint: you can use the result from the last question to help you prove this one).

Programming Skills - Matlab [5pts]

Sampling from a distribution.

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

2. How does the scatter plot change if the mean is $(1, 1)^T$ (and identity covariance)? Submit your plot.
3. How does the scatter plot change if the covariance matrix is $\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ (and zero mean)? Submit your plot.
4. How does the scatter plot change if the covariance matrix is changed to $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ (and zero mean)? Submit your plot.
5. How does the scatter plot change if the covariance matrix is changed to $\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$ (and zero mean)? Submit your plot.