Homework 0: Mathematical Background for Machine Learning

10-401 Machine Learning
Due 5 p.m. Wednesday, January 24, 2018

The goal of this homework is to help you refresh the mathematical background needed to take this class. 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 don't need to be an expert in all these areas, but you will need to be conversant in each, and to understand:

- 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.
- Basic calculus (at the level of a first undergraduate course). For example, we rely on you being able to take derivatives. During the class we will sometimes 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.

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.

Please see the piazza post for useful resources for brushing up on, and filling in this background.

Instructions

- **Submit your homework** to Autolab by 5 p.m. Friday, January 24, 2015.
- **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 refresh the background needed for this class, and to fill in any gaps you may have.

Minimum Background Test [80 Points]

Vectors and Matrices [20 Points]

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

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

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 $\mathbf{y^T z}$)
2. What is the product $\mathbf{Xy}$?
3. Is $\mathbf{X}$ invertible? If so, give the inverse, and if no, explain why not.
4. What is the rank of $\mathbf{X}$?

Calculus [20 Points]

1. If $y = x^3 + x - 5$ then what is the derivative of y with respect to x?
2. If $y = x \sin(z) e^{-x}$ then what is the partial derivative of y with respect to x?

Probability and Statistics [20 Points]

Consider a sample of data $S = \{1, 1, 0, 1, 0\}$ 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 coin with an equal probability of heads and tails (i.e. the probability distribution is $p(x = 1) = 0.5$, $p(x = 0) = 0.5$).
4. Note that the probability of this data sample would be greater if the value of $p(x = 1)$ was not 0.5, 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 over variables y and z, where y takes a value from the set {a,b,c}, and z takes a value from the set {T,F}:

		y		
		a	b	c
z	T	0.2	0.1	0.2
	F	0.05	0.15	0.3

- What is $p(\text{z} = \text{T} \quad \text{AND} \quad \text{y} = \text{b})$?
- What is $p(\text{z} = \text{T}|\text{y} = \text{b})$?

Big-O Notation [20 Points]

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. $f(n) = \ln(n)$, $g(n) = \lg(n)$.
   Note that ln denotes log to the base e and lg denotes log to the base 2.
2. $f(n) = 3^n$, $g(n) = n^{10}$
3. $f(n) = 3^n$, $g(n) = 2^n$

Medium Background Test [20 Points]

Algorithms [5 Points]

Divide and Conquer: Assume that you are given an array with n elements all entries equal either to 0 or +1 such that all 0 entries appear before +1 entries. You need to find the index where the transition happens, i.e. you need to report the index with the last occurrence of 0. Give an algorithm that runs in time $O(\log n)$. Explain your algorithm in words, describe why the algorithm is correct, and justify its running time.

Probability and Random Variables [5 Points]

Probability
State true or false. Here A^c denotes complement of the event A.
(a) $P(A \cup B) = P(A \cap (B \cap A^c))$
(b) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(c) $P(A) = P(A \cap B) + P(A^c \cap B)$
(d) $P(A|B) = P(B|A)$
(e) $P(A_1 \cap A_2 \cap A_3) = P(A_3|(A_2 \cap A_1))P(A_2|A_1)P(A_1)$

Discrete and Continuous Distributions
Match the distribution name to its formula.

Multivariate Gaussian	$p^x(1-p)^{1-x}$
Bernoulli	$\frac{1}{b-a}$ when $a \leq x \leq b; 0$ otherwise
Uniform	$\binom{n}{x}p^x(1-p)^{n-x}$
Binomial	$\frac{1}{\sqrt{(2\pi)^d\|\Sigma\|}} \exp\left(-\frac{1}{2}-(\mathbf{x}-\mu)^\top\Sigma^{-1}(\mathbf{x}-\mu)\right)$

Mean, Variance and Entropy
(a) What is the mean, variance and entropy 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 variance of $x+2$?

Law of Large Numbers and Central Limit Theorem
Provide one line justifications.
(a) If a die is rolled 6000 times, the number of times 3 shows up is close to 1000.
(b) If a fair coin is tossed n times and $\bar{X}$ denotes the average number of heads, then distribution of $\bar{X}$ satisfies

$$\sqrt{n}(\bar{X} - 1/2)] \overset{n\to\infty}{\to} \mathcal{N}(0, 1/4)$$

Linear Algebra [5 Points]

Vector norms

Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with following norms:
(a) $\|\mathbf{x}\|_2 \leq 1$ (Recall $\|x\|_2 = \sqrt{\sum_i x_i^2}$)
(b) $\|\mathbf{x}\|_0 \leq 1$ (Recall $\|x\|_0 = \sum_{i:x_i \neq 0} 1$)
(c) $\|\mathbf{x}\|_\infty \leq 1$ (Recall $\|x\|_\infty = \max_i |x_i|$)

Geometry

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

Programming skills - your favorite language, e.g., MATLAB/R/C [5 Points]

Sampling from a distribution

(a) Draw 100 samples $\mathbf{x} = [x_1 \; x_2]$ from a 2-dimensional Gaussian distribution with mean $[0, 0]$ and identity covariance matrix i.e. $p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right)$, and make a scatter plot (x_1 vs. x_2).
(b) How does the scatter plot change if the mean is $[-1, 1]$?
(c) How does the scatter plot change if you double the variance of each component?
(d) How does the scatter plot change if the covariance matrix is changed to the following?

$$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

(e) How does the scatter plot change if the covariance matrix is changed to the following?

$$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}$$