10-601/10-701 Pre-requisites

Although many students find the machine learning class to be very rewarding, the class does 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). 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). 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.

- Basic probability and statistics (at the level of a first undergraduate course). For example, we assume you already 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.

To help you self-evaluate whether you have the background to succeed in the class, below we have produced a simple self-evaluation test. For each of these mathematical 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 good 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.

Some useful resources for brushing up on, and filling in this background include:
1. Probability review:
   http://www.cs.cmu.edu/~aarti/Class/10701/recitation/prob_review.pdf
2. Linear Algebra review:
   http://www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf
   http://www.cs.cmu.edu/~aarti/Class/10701/recitation/LinearAlgebra_Matlab_Review.ppt
   Book: Gilbert Strang. Linear Algebra and its Applications. HBJ Publishers.

Necessary minimum background test
(this should take 10-15 minutes, if you know the material)

1. Multivariate calculus

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

2. Vectors and matrices

Consider the matrix **X** and the vector **y** below
$$\mathbf{X} = \begin{bmatrix} 2 & 4 \\ 1 & 3 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

What is the product **Xy** ?
Is **X** invertible? If so, give the inverse, if not explain why not.
What is the rank of **X** ?

3. Probability and statistics

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}

What is the sample mean for this data ?

What is the sample variance ?

What is the probability of observing this data assuming that a coin with an equal probability of heads and tails was used (i.e., by the probability distribution p(x=1)=0.5, p(x=0)=0.5).

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?]

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

Modest Background Test

1 Probability and Random Variables

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 $\quad p^x(1-p)^{1-x}$

Exponential $\quad \frac{1}{b-a}$ when $a \le x \le b$; 0 otherwise

Uniform $\quad \binom{n}{x}p^x(1-p)^{n-x}$

Bernoulli $\quad \lambda e^{-\lambda x}$ when $x \ge 0$; 0 otherwise

Binomial $\quad \frac{1}{\sqrt{(2\pi)^d|\Sigma|}} \exp\left(-\frac{1}{2}-(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)\right)$

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 $x + 2$?

Mutual and Conditional Independence
(a) If X and Y are independent random variables, show that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.
(b) 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?

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

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

Reading material:
http://www.cs.cmu.edu/~aarti/Class/10701/recitation/prob_review.pdf

2   Linear Algebra

Vector norms
Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with following norms: (a) $\|\mathbf{x}\|_2 \le 1$ (Recall $\|x\|_2 = \sqrt{\sum_i x_i^2}$)
(b) $\|\mathbf{x}\|_0 \le 1$ (Recall $\|x\|_0 = \sum_{i:x_i \ne 0} 1$)
(c) $\|\mathbf{x}\|_1 \le 1$ (Recall $\|x\|_1 = \sum_i |x_i|$)
(d) $\|\mathbf{x}\|_\infty \le 1$ (Recall $\|x\|_\infty = \max_i x_i$)

Matrix Decompositions and Rank
a) Give the definition of the eigenvalues and the eigenvectors of a square matrix.
b) Find the eigenvalues and eigenvectors of

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

c) Show that the eigenvalues of $\mathbf{A}^k$ are $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$, the kth powers of the eigenvalues of matrix $\mathbf{A}$, and that each eigenvector of $\mathbf{A}$ is still an eigenvector of $\mathbf{A}^k$.

Vector and Matrix Calculus
(a) What is the first derivative of $\mathbf{a}^T\mathbf{x}$ with respect to $\mathbf{x}$?
(b) What is the first derivative of $\mathbf{x}^\top \mathbf{A}\mathbf{x}$ with respect to $\mathbf{x}$? What is the second derivative?

Reading Material:
http://www.cs.cmu.edu/~aarti/Class/10701/recitation/LinearAlgebra_Matlab_Review.ppt
http://www.cs.cmu.edu/~zkolter/course/15-884/linalg-review.pdf
Wikipedia: http://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors
Gilbert Strang. Linear Algebra and its Applications, Ch 5. HBJ Publishers.

3   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 $\frac{b}{\|\mathbf{w}\|}$.

4 Programming skills - MATLAB/R/C

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?

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

Eigendecomposition

Compute the eigenvector corresponding to the largest eigenvalue of the following matrix.

$$\begin{pmatrix} 1 & 0 \\ 1 & 3 \end{pmatrix}$$

Reading material:

Matlab tutorial - http://www.math.mtu.edu/~msgocken/intro/intro.pdf

R tutorial - http://math.illinoisstate.edu/dhkim/rstuff/rtutor.html