Math preparation materials for the MS in Data Science program
and prep for exemption exam for DS5020 Linear Algebra and Probability for DS

The following materials are provided to help you prepare for entry into the MS in Data Science program. You will be tested on this material during the first week of your starting semester. The purpose of the exam is to evaluate your knowledge and proficiency in required mathematics to be successful in the Data Science program.

Students should be comfortable with the material provided in the following documents/websites:

- Probability and Statistics (review material in section 5 in these [notes](#))

- Linear Algebra - https://course.ccs.neu.edu/ds5230f18/assets/pdf/cs229-linalg.pdf

- Random Vectors and Matrices (review material in provided [here](#))

- Unitary Matrices, Change of Basis (review material in sections 5.1-5.5 in these [notes](#))

In you need more background in any of these topics, please refer to the textbooks in the DS5020 Study Guide. The following list of alternative study materials can also be useful:

· Probability Theory Review (A. Maleki and T. Do, Stanford). Free online.
   http://cs229.stanford.edu/section/cs229-prob.pdf
· Linear Algebra Tutorial (C.T. Abdallah, Penn). Free online.
   http://www.seas.upenn.edu/~jadbabai/ESE504/LAreview.pdf
· Linear Algebra Review and Reference (Z. Kolter and C. Do, Stanford). Free online.
   http://cs229.stanford.edu/section/cs229-linalg.pdf

Additionally, here are some links to other materials to take a look at:

- [Probability Review (David Blei, Princeton)](#)
- [Probability Theory Review (Arian Maleki and Tom Do, Stanford)](#)
- [Linear Algebra Tutorial (C.T. Abdallah, Penn)](#)
- [Linear Algebra Review and Reference (Zico Kolter and Chuong Do, Stanford)](#)
- [Probability, Linear Algebra, and Differentiation (Iain Murray)](#)
- [Statistical Data Mining Tutorials (Andrew Moore, Google/CMU)](#)
- [Theoretical CS Cheat Sheet (Princeton)](#)

The exam may include, but is not limited to, questions such as those below.

1. You flip a fair coin 10 times. You get 4 heads and 6 tails. What is the probability of getting a head on your next coin flip?

2. Given the distribution {-2, -1, 3, 3, 5, 6, 8, 10, 11}, what are the mean, median and standard deviation of the sample?

3. What is the value of the limit: $\lim$ $\underline{(x^2 - x - 2)}$

$$x \to 2 \quad (x^2 - 2x)$$

a) 0   b) -1   c) 2   d) infinity   e) the limit does not exist

4. We are given that the probability density function (pdf) of a continuous random variable X is:

$$p(x) = \begin{cases} 4x & 0 \le x \le 0.5 \\ -4x + 4 & 0.5 \le x \le 1 \end{cases}$$

   What is the equation for the corresponding cumulative density function (cdf) *P(x)?*

5. Consider the following system of equations:

   $2x_1 + x_2 + x_3 = 3$
   $4x_1 + 2x_3 = 10$
   $2x_1 + 2x_2 = -2$

   Solve for x_1, x_2, and x_3.

6. Consider the following matrix:

1	2	3
1	4	3
1	3	4

   (a.)   Is the matrix invertible?
   (b.)   What is the rank of the matrix?

7. The eigenvalues of the matrix $A = \begin{matrix} 3 & 6 \\ 1 & 4 \end{matrix}$

   are $\lambda = 6$ and $\lambda = 1$.   Which of the following is an eigenvector for $\lambda = 1$?

   a. $\begin{matrix} 0 \\ 1 \end{matrix}$ b. $\begin{matrix} -3 \\ 1 \end{matrix}$ c. $\begin{matrix} 3 \\ 1 \end{matrix}$ d. $\begin{matrix} 1 \\ 0 \end{matrix}$

8. Find the determinant of A. $A = \begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 4 & 4 \\ 1 & 4 & 9 & 9 \\ 1 & 4 & 9 & 16 \end{matrix}$

9. The weatherman has predicted rain tomorrow. In recent year, it has rained only 73 days each year. When it actually rains, the weatherman correctly forecasts rain 70% of the time. When it does rain, he incorrectly forcasts rain 30% of the time. What is the probability that it will rain tomorrow?
   Hint: Bayes Rule formula is P (X|E) = (P(E|X) * P(X))/P(E