CSCI 5521: Introduction to Machine Learning (Fall 2017)[1]

Homework 0

Due date: Monday, Sep 18th, 11:55pm

1. (a) Which of the following courses have you taken?
   i. Artificial Intelligence II
   ii. Introduction to Data Mining
   iii. (Advanced) Algorithms

   (b) Have you taken any course on Probability/Statistics? If yes, please write down the course title (not number) and department. [2]

   (c) Have you taken any course on Numerical Methods/Linear Algebra? If yes, please write down the course title (not number) and department.[2]

2. (a) Given a full rank matrix $A \in \mathbb{R}^{m \times n}$ where $m > n$ and $B \in \mathbb{R}^{m \times k}$, show how to solve the following system of equations:

   $$AX = B$$

   It is sufficient to outline the method at a high level.

   (b) What happens if A is not full rank? Briefly explain your answer.

3. Given two vectors a and b, explain one way to calculate the distance between the vectors (you can give an example). Show the same for two matrices A and B.
   Note: your distance should be one scalar number that represents the distance between the two vectors/matrices.

4. Given two independent random variables A and B, which of the following expressions are true (you can choose more than one)?

   (a) $P(A, B) = P(A) \times P(B)$

   (b) $P(B|A) = P(B)$

   (c) $P(A) + P(B) = 1$

   (d) $P(A, B) = P(A) + P(B)$

[1] Instructor: Catherine Qi Zhao (qzhao@umn.edu). TAs: Raphael Petegrosso (peteg001@umn.edu); Xinyan Li (lixx1166@umn.edu).

[2] For example, for the current course, the information will be: 'Introduction to Machine Learning, Department of Computer Science & Engineering.' It is ok if you have taken the course at a different University, we do not need that information.

5. (a) Let D be a random variable for a given disease, assume that the probability a person has the disease is 0.1. Based on this information, researchers developed a new method to say if a person has the disease: for each 10 people that do the test, they randomly report that 1 of them has the disease. Will the method correctly identify if the person has the disease? Briefly explain your answer.

   (b) Another group of researchers developed a new blood test to identify the same disease. The test result is given by a random variable X, with sensitivity and specificity given by 0.8 and 0.9, respectively (that means $p(X = 1|D = 1) = 0.8$ and $p(X = 0|D = 0) = 0.9$). If a patient did the blood test and the result is positive, what is the probability that the person has the disease?
   Hint: you might want to use the Bayes Rule: $p(b|a) = \frac{p(a|b)p(b)}{p(a)}$

Submission

- **Things to submit:**
  1. hw0_sol.pdf: a document containing all your answers.
- **Submit**: All material must be submitted electronically via Moodle.