CIS 419/519: Applied Machine Learning | Spring 2020

Homework 0

Handed Out: January 22 | *Due: January 27*

- The goal of HW0 is to give you an idea of the level of mathematical knowledge and maturity expected in this course. You should have seen all the material before; the goal of this homework is to encourage you to go back to some of the material and refresh your memory.
- Most of these problems require you to be familiar with the notation and concepts. Beyond that, you should be able to reason from the definitions and do some algebraic manipulation. None of these problems require deep understanding of the material.
- Feel free to go back to some of your text books and the web when solving it. You should, however, write down your own solution to these problems.
- In the past, we have seen many students start the course who did not have an adequate background to complete the course. Most of these students ended up dropping out halfway through, or receiving a poor grade. It is for this reason that we have designed HW0 to test this background knowledge (with assistance from textbooks and other sources, of course). **If you are not familiar with more than 30% of the material in this problem set, or find it difficult to solve at least two-thirds of it in a few hours, you are missing the prerequisites, and you will find the class too difficult. In this case, I highly recommend that you drop the class, take the required classes first, and then come back to take this class.**
- While we encourage discussion within and outside the class, in the specific case of HW0, you are on your own. It's a test of your level of readiness to take this course, so please work on it independently.
- Please use Piazza if you have questions about the homework. For all future assignments, we encourage you to come to the Professor and the TAs office hours and recitations for assistance. However, for this specific assignment, you must complete it entirely on your own, without consulting course staff for assistance.
- Please try to keep your solutions brief and clear.
- You are required to format your solutions properly using LaTeX; handwritten solutions are not permitted. You will submit your solution manuscript as a single PDF file.
- The homework is due at 11:59 PM on the due date. We will be using Gradescope for collecting the homework assignments. Please submit your solution manuscript as a PDF file via Gradescope. Hard copy solutions will not be accepted. Post on Piazza and contact the TAs if you are having technical difficulties in submitting the assignment.
- **If you are on the waiting list**, we recommend that you still work on the assignment so that you can determine if you want to stay on the waiting list.

1 Multiple Choice & Written Questions

Note: You do not need to show work for multiple choice questions. Be sure to download and use our hw0_answer_template.tex when formatting your answer in LaTeX.

1. [Calculus] Let $f(x, y) = 3x^2 + y^2 - xy - 11x$

   a. (5pts) Find $\frac{\partial f}{\partial x}$, the partial derivative of f with respect to x. Find $\frac{\partial f}{\partial y}$. Select one of the below four options.

   A) $\frac{\partial f}{\partial x} = 6x - 6y - 10, \frac{\partial f}{\partial y} = y - x$
   B) $\frac{\partial f}{\partial x} = x - y - 11, \frac{\partial f}{\partial y} = 2y - x$
   C) $\frac{\partial f}{\partial x} = 6x - y - 11, \frac{\partial f}{\partial y} = 2y - x$
   D) $\frac{\partial f}{\partial x} = x - 6y - 11, \frac{\partial f}{\partial y} = 2y - x$

   b. (5pts) Find $(x, y) \in \mathbb{R}^2$ that minimizes f. Select one of the below four options.

   A) $(2, 1)$
   B) $(1, 1)$
   C) $(2, 2)$
   D) $(1, 2)$

2. [Probability] Assume that the probability of obtaining heads when tossing a coin is λ.

   a. (5pts) What is the probability of obtaining the first head on the $(k + 1)$-th toss? Select one of the below four options.

   A) $(1 - \lambda)^k \lambda^k$
   B) $(1 - \lambda)\lambda$
   C) $(1 - \lambda)\lambda^k$
   D) $(1 - \lambda)^k \lambda$

   b. (5pts) What is the expected number of tosses needed to get the first head? Select one of the below four options.

   A) $\frac{1}{\lambda}$
   B) $\frac{1}{\lambda^2}$
   C) $\frac{1-\lambda}{\lambda}$
   D) $\frac{1}{\lambda^3}$

3. [Probability] John is a great fortune teller. Assume that we know three facts:

   - If John tells you that a lottery ticket will win, it will win with probability 0.99.
   - If John tells you that a lottery ticket will not win, it will not win with probability 0.99999.
   - Seeing a ticket, John predicts with probability 10^{-5} that a ticket is a winning ticket, implying, with probability $1 - 10^{-5}$, he predicts that a ticket won't win.

   With this information, please answer the two questions below:

a. (5pts) Given a ticket, what is the probability that it wins? Select one of the below four options.

A) $\approx 1.99 \times 10^{-4}$
B) $\approx 2.667 \times 10^{-5}$
C) $\approx 2.667 \times 10^{-3}$
D) $\approx 1.99 \times 10^{-5}$

b. (5pts) What is the probability that John correctly predicts a winning ticket? Select one of the below four options.

A) 0.398
B) 0.523
C) 0.497
D) 0.667

4. [Probability] The variance of a random variable X is defined as

$$Var(X) = E[(X - E[X])^2] \ .$$

a. (5pts) If $E[X] = 0$ and $E[X^2] = 1$, what is the variance of X? If $Y = a + bX$, where a and b are constants, what is the variance of Y? Select one of the below four options.

A) $Var(X) = \frac{1}{\sqrt{2}}, Var(Y) = a^2 + b^2$
B) $Var(X) = 1, Var(Y) = b^2$
C) $Var(X) = \sqrt{2}, Var(Y) = a^2$
D) $Var(X) = 2\sqrt{2}, Var(Y) = b^2$

b. (10pts) Prove that $Var(X) = E[X^2] - E[X]^2$. Please show your work in detail.

2 Python Programming Questions

Download the following iPython notebook: iPython notebook for HW 0. You will solve all of the programming problems using that iPython notebook.

5. [Programming] (25pts) You're trapped in the bottom-left corner of a $w \times h$ escape room, and need to cross to the door in the upper-right corner to escape. But, the door will only unlock if you follow the correct path. Fortunately, you have directions for the path to follow across the grid cells on the floor, but the directions are incomplete. The partial path is represented as a string composed of the following characters {'U', 'D', 'L', 'R', '?'}, where each character represents a direction to take along the grid: U = up, D = down, L = left, R = right. The question marks represent unknown steps that you must figure out. You know that the path may not cross the same grid cell twice.

Write a function `CompletePath(s, w, h)` that takes in three arguments: a string `s` representing the partial path, and two integers `w` and `h` specifying the room size. Your

function should return a string that represents the completed path, with all question marks filled in with the correct directions to take you from the bottom-left corner of the grid to the upper-right corner. If no path is possible, your function should return `None`. Your function should properly error-check all arguments.

For example: in a 5×5 room, if `s = "URD?RUU?UR"` then your function should return the completed path to exit the room, `"URDRRUUUUR"`.

Here are a few sample test cases:
`CompletePath("?RDRR?UUUR", 5, 5) = "URDRRUUUUR"`
`CompletePath("UURDD?UUR?RR", 6, 4) = "UURDDRUURURR"`

Complete this function in the iPython notebook.

6. [Pandas Data Manipulation] (30pts) In this problem, we will write three data manipulation functions for Pandas dataframes:

   - `getMissingRatio(inputDf)` – Determines the missing data percentage of each feature.
   - `convertToBinary(inputDf, feature)` – Converts a binary categorical feature into a numerical 0-1 representation and replaces it in the dataframe.
   - `addDummyFeatures(inputDf, feature)` – Create a one-hot-encoded version of a categorical feature and append it to the existing dataframe. For background on one-hot encoding, please see below.

   Complete these functions in the iPython notebook.

   Background on one-hot encoding: Some machine learning implementations, such as sklearn's decision tree classifier, cannot handle categorical features that aren't encoded numerically. One way of encoding categorical features is using a "one-hot encoding". If a categorical feature takes on three values, e.g. color $\in$ {red, green, blue}, then the one-hot encoding would convert each of those values into a binary feature: one feature would be for whether the value is red, another for whether the value is green, and another for whether the value is blue. Notice that only one of those binary features is positive ("hot") for each instance, hence the name. Consider the small three feature data set below:

color	**weight**	**label**
red	2.5	yes
green	2.2	no
blue	2.8	yes
green	3.2	no
blue	3.7	no

$\rightarrow$

color_red	**color_green**	**color_blue**	**weight**	**label**
1	0	0	2.5	yes
0	1	0	2.2	no
0	0	1	2.8	yes
0	1	0	3.2	no
0	0	1	3.7	no