Homework 0: Preliminaries

CSCI 375: NLP / Williams College / Fall 2024

The objectives of this homework are to: (1) help the instructors get to know you via a course survey, (2) successfully set-up the programming environment you'll use throughout the semester and (3) help you review Jupyter Notebooks and Python, and (4) ensure that you are up to speed on the necessary mathematical and statistical prerequisites.

Grading. This homework is intended to help you assess where you are in the course and will not be graded or count towards the homework portion of your final grade.

1 Course survey

Fill out this course survey to help the instructor and TAs get to know you better and set you up for success in this class.

2 Computing

Navigate to this Github page and follow the instructions given.

There are three parts:

1. *Required:* Programming environment set-up
2. *Optional but recommended:* Jupyter notebook tutorial
3. *Optional but recommended:* Python tutorial

There is no deliverable for this part of the assignment. However, all future assignments require that you have set-up the programming environment successfully and that you understand Python and Jupyter Notebooks. So please complete these steps accordingly.

3 Math prerequisites take-home quiz

The math prerequisites take-home quiz is on the next page. This part will be graded on effort—if you attempt the quiz (no matter your score) you will receive 100%. However, we ask that you take this quiz seriously as it will help prepare you for success in the class.

When you finish, submit your write-up to **Gradescope**.

Math prerequisites take-home quiz

Deliverables. Please write-up your answers and submit them via PDF to Gradescope.

Instructions.

For this take-home quiz, please record the following steps in your PDF submission:

1. *First pass.* Sit uninterrupted for (maximum) 1 hour and answer the questions below. Pretend this is a quiz and do not access outside help (e.g. the internet or a friend). Feel free to skip any questions you do not know how to answer.
2. *Reflection.* Once you've finished, write a short (1-3 sentence) reflection on your first pass. Which concepts do you feel comfortable on? Which concepts do you need to review more? Can you identify the underlying principle that can help you answer the question? We highly recommend reading the "Eisenstein. Appendix A: Probability" **from the course reading packet**.
3. *Second pass.* In your second pass through questions you didn't understand, try to answer the question again, this time with access to resources.

Please be honest with your first pass, reflection, and second pass. This can help you and the instructors identify where we can help you so you can succeed with the more advanced materials in subsequent assignments.

We will provide solutions once everyone has turned in their take-home quizes so you can self-assess.

Problems.

1. A and B are discrete random variables. A can take 4 possible values and B can take 7 possible values. How many possible outcomes does the joint distribution $P(A, B)$ define probabilities for?
2. Suppose we have a sequence of n binary variables $A_1, A_2, \ldots, A_n$. How many possible outcomes does the joint distribution $P(A_1, A_2, \ldots, A_n)$ define probabilities for? *Hint:* Your answer should be in terms of n.
3. Suppose we have three random variables, A, B, and C. Which of the following statements are *always true*?
   a. $P(A|B) = P(B|A)$
   b. $P(A, B) = P(A|B)P(B)$
   c. $P(A, B) = P(A)P(B)$
   d. $P(A|B) = P(A)$
   e. $P(A, B, C) = P(A)P(B)P(C)$
   f. $P(A, B, C) = P(A)P(B|A)P(C|A, B)$
   g. $P(A) = \sum_{b \in \text{domain}(B)} P(A, B = b)$
   h. $P(A) = \sum_{b \in \text{domain}(B)} P(A|B = b)P(B = b)$
4. Assuming A, B and C are independent, which of the statements a-h from the question above are true?
5. Let p be a probability, i.e. $0 \leq p \leq 1$. For what values of p is $\log(p)$ defined? What are the range of possible values for $\log(p)$? Here we are defining $\log(\cdot)$ as the natural logarithm (base e).
6. Let p and q both be probabilities, i.e. $0 \leq p \leq 1$ and $0 \leq q \leq 1$.

a. What is the range of possible values for p/q?

b. What is the range of possible values for $\log(p/q)$?

7. What is $\frac{\partial}{\partial x}\log(2x^3)$?

8. What is $\frac{\partial}{\partial u}\frac{1}{1+e^{-u}}$?

Credit: Some of these questions were adapted from Brendan O'Connor's HW0, https://people.cs.umass.edu/~brenocon/cs490a_f21/hw0.pdf.