DSC 40A Fall 2024 - Group Work Session 6 - ASYNCHRONOUS

due Wednesday, Nov 13th at 11:59PM

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **tag all group members** so everyone gets credit.

This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

1 Gradient Descent

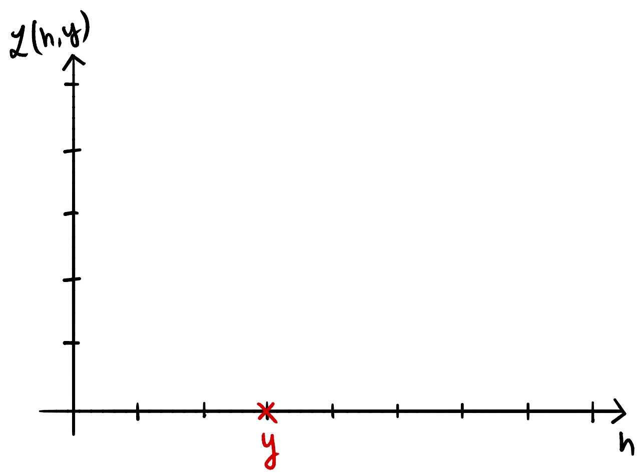
Gradient descent is an algorithm used to minimize differentiable functions. In this class, we will primarily use it to minimize empirical risk, though its use is much more broad. In this problem, we'll explore a new loss function and see how our initial prediction impacts how much our prediction changes with one iteration.

Problem 1.

Consider a new loss function,

$$L(h, y) = \begin{cases} (h-y)^2, & h < y \\ (h-y)^3 & h \geq y \end{cases}.$$

a) Fix an arbitrary value of y. On the axes below, draw the graph of $L(h, y)$ as a function of h.



b) For any data set $y_1, \ldots, y_n$, the empirical risk, $R(h) = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} L(h, y_i)$ will be differentiable and convex (also known as concave up). This means gradient descent is guaranteed to be able to find its minimum value. It might take many iterations or only a few. We won't do the whole algorithm, just one iteration for practice.

Suppose our data set is $\{2, 3, 6, 10\}$. Perform one iteration of gradient descent by hand on the empirical risk function $R(h)$ for this data set, starting with an initial prediction of $h_0 = 3$ and using a step size of $\alpha = \frac{1}{10}$. Calculate h_1, the prediction after the first iteration.

c) For the same data set $\{2, 3, 6, 10\}$, suppose instead we did one iteration of gradient descent on $R(h)$ but starting at a different initial prediction, $h_0 = 7$. Using the same step size of $\alpha = \frac{1}{10}$, calculate h_1, the prediction after the first iteration, for this new starting point.

d) Compare your answers to the previous parts and notice that the prediction moves only a little bit when we start at $h_0 = 3$, but it moves a lot when we start at $h_0 = 7$. Can you explain why that happens by looking at the loss function we're using?

2 Probability

Most probability questions can be solved by applying one of the basic probability rules in the right way. Sometimes, some cleverness is needed to define the right sample space or the right events. There are often many ways to solve the same problem, some easier than others. It's really useful to learn multiple ways of doing the same problem, which will help you develop your problem-solving skills.

Here are the basic probability rules you'll need to use to solve the questions that follow.

Addition Rule:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Multiplication Rule:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A)$$

Complement Rule:

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$$

Conditional Probability:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Problem 2.

A bitstring is a sequence of 0s and 1s. For example, 0110100 is a bitstring of length 7.

Suppose that we generate a bitstring of length 4 such that each digit is equally likely to be a 0 or 1.

a) What is the probability that the bitstring is 1111?

b) What's the probability that the bitstring contains at least one 0 and one 1?

c) What is the probability that a bitstring has more 0s than 1s?

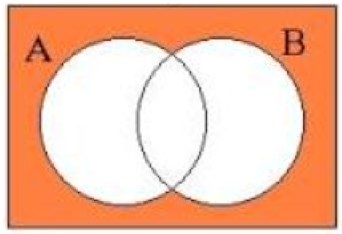
d) What is the probability that a bitstring has more 0s than 1s, if we know that the first bit is a 0?

e) Suppose now that you generate two bitstrings and look at one of them. You see that this bitstring has more 0s than 1s. What is the probability that in total, for both strings together, there are more 0s than 1s?

Problem 3.

Let A and B be two independent events in the sample space S. Show that $\bar{A}$ and $\bar{B}$ must be independent of one another.

You may use the fact that $P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B)$, which should be apparent from the Venn diagram below.



Problem 4.

Suppose you have 6 pairs of socks in your sock drawer, each in a different pattern. It is still dark out in the morning when you get dressed, so you randomly pull one sock at a time out of the drawer, until you have removed two matching socks. What is the probability that you pull out exactly 5 socks from your sock drawer in the morning?

Problem 5.

You're listening to a YouTube playlist of the 14 songs in Oscar Peterson's album, *Solo*. Suppose you're listening on shuffle, and each time a new song starts, it's equally likely to be any of the 14 songs on the playlist, regardless of which songs have been played so far. How many songs must you listen to so that the probability of hearing "Mirage" is at least 75%?

Problem 6.

Suppose we scramble the 26 letters of the alphabet in a random order so that each rearrangement is equally likely. What is the probability that the letters ABC wind up next to each other in that order?