CS221 Summer 2023: Artificial Intelligence: Principles and Techniques

Homework 1: Foundations

SUNet ID: alebarro
Name: Alessandro Barro
5Collaborators:

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Welcome to your first CS221 assignment! The goal of this assignment is to sharpen your math, programming, and ethical analysis skills needed for this class. If you meet the prerequisites, you should find these problems relatively innocuous. Some of these problems will occur again as subproblems of later homeworks, so make sure you know how to do them. If you're unsure about them or need a refresher, we recommend going through our prerequisites module or other resources on the Internet, or coming to office hours.

Before you get started, please read the Assignments section on the course website thoroughly.

Problem 1: Optimization and probability

In this class, we will cast a lot of AI problems as optimization problems, that is, finding the best solution in a rigorous mathematical sense. At the same time, we must be adroit at coping with uncertainty in the world, and for that, we appeal to tools from probability.

a. Let $x_1, \dots, x_n$ be real numbers representing positions on a number line. Let $w_1, \dots, w_n$ be positive real numbers representing the importance of each of these positions. Consider the quadratic function: $f(\theta) = \sum_{i=1}^{n} w_i(\theta - x_i)^2$. Note that θ here is a scalar. What value of θ minimizes $f(\theta)$? Show that the optimum you find is indeed a minimum. What problematic issues could arise if some of the w_i's are negative?
[**NOTE:** You can think about this problem as trying to find the point θ that's not too far away from the x_i's. Over time, hopefully you'll appreciate how nice quadratic functions are to minimize.]

[**What we expect:** An expression for the value of θ that minimizes $f(\theta)$ and how you got it. A short calculation/argument to show that it is a minimum. 1-2 sentences describing a problem that could arise if some of the w_i's are negative.]

Your Solution: In order to find the value of θ that minimizes $f(\theta)$, we derive the function with respect to θ itself

$$\frac{\partial f(\theta)}{\partial \theta} = \sum_{i=1}^{n} 2w_i(\theta - x_i)$$

Now, let's set the expression equal to 0 and manipulate it a little bit

$$\sum_{i=1}^{n} 2w_i(\theta - x_i) = 0$$
$$\sum_{i=1}^{n} w_i\theta - \sum_{i=1}^{n} w_i x_i = 0$$

We finally obtain

$$\theta = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} \; *$$

Which corresponds to the stationary point of the quadratic function $f(\theta)$. In order to verify if this one is a minimum, we should take a look at the second derivative

$$\frac{\partial^2 f(\theta)}{\partial^2 \theta} = 2\sum_{i=1}^{n} w_i > 0$$

Which, since $w_i > 0$ by hypothesis, is definitely greater than 0 and the value found is indeed a minimum.

If some of the w_i are negative, the statement we just made could be not true, and $f(\theta)$ could be concave instead of convex, and so impossible to minimize. Also note that $\sum_{i=1}^{n} w_i \neq 0$, as a condition of existence in *.

b. In this class, we will frequently encounter operators such as sum, min, and max. Let's explore what happens if we switch the order of these operators.

Let $f(\mathbf{x}) = \min_{s \in [-1,1]} \sum_{i=1}^{d} s x_i$ and $g(\mathbf{x}) = \sum_{i=1}^{d} \min_{s_i \in [-1,1]} s_i x_i$, where $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ is a real vector and $[-1, 1]$ means the closed interval from -1 to 1. Which of $f(\mathbf{x}) \leq g(\mathbf{x})$, $f(\mathbf{x}) = g(\mathbf{x})$, or $f(\mathbf{x}) \geq g(\mathbf{x})$ is true for all $\mathbf{x}$? Prove it.
[**HINT:** You may find it helpful to refactor the expressions so that they are minimizing the same quantity over different sized sets.]

[**What we expect:** A short (3-5) line/sentence proof. You should use mathematical notation in your proof, but can also make your argument in words.]

Your Solution: Let's put the focus on the s and s_i terms in the expressions of $f(x)$ and $g(x)$. In $f(x)$ we can note that a unique value s is applied to the sum $\sum_{i=1}^{d} x_i$. In $g(x)$, s_i is applied to every addend x_i distinctively. If we look forward into minimizing those functions, we can safely say that $f(x) \geq g(x)$ for all x, based on our assumptions.

c. Suppose you repeatedly roll a fair six-sided die until you roll a 1 or 2 (and then you stop). Every time you roll a 3, you win a points, and every time you roll a 4, you lose b points. You do not win or lose any points if you roll a 5 or 6. What is the expected number of points (as a function of a and b) you will have when you stop?
[**HINT:** You will find it helpful to define a recurrence. If you define V as the expected number of points you get from playing the game, what happens if you roll a 3? You win a points and then get to play again. What about the other cases? Can you write this as a recurrence?]

[**What we expect:** A recurrence to represent the problem and the resulting expression from solving the recurrence (no more than 1-2 lines)]

Your Solution: Each outcome has $\frac{1}{6}$ of probability and we can formalize it as follows $V = \frac{1}{6}(0 + 0 + a + V - b + V + V + V)$. Solving by V we obtain $V = \frac{a-b}{2}$ (the expected number of points).

d. Suppose the probability of a coin turning up heads is p (where $0 < p < 1$), and we flip it 4 times and get $\{\mathrm{T}, \mathrm{H}, \mathrm{H}, \mathrm{H}\}$. We know the probability (likelihood) of obtaining this sequence is $L(p) = (1-p)ppp = p^3(1-p)$. What value of p maximizes $L(p)$? Prove/Show that this value of p maximizes $L(p)$. What is an intuitive interpretation of this value of p?
[**HINT:** Consider taking the derivative of $\log L(p)$. You can also directly take the derivative of $L(p)$, but it is cleaner and more natural to differentiate $\log L(p)$. You can verify for yourself that the value of p which maximizes $\log L(p)$ must also maximize $L(p)$ (you are not required to prove this in your solution).]

[**What we expect:** The value of p that maximizes $L(p)$ and the work/calculation used to solve for it. Note that you must prove/show that it is a maximum. A 1-sentence intuitive interpretation of the value of p.]

Your Solution: To find the value of p that maximizes $L(p)$, that we will call $\ddot{p}$, we first apply *log* to the function and then derive it

$$\frac{d \log L(p)}{dp} = \frac{4p-3}{(p-1)p}$$

Let's set now $\frac{d \log L(p)}{dp} = 0$ to find $\ddot{p}$

$$\frac{4p-3}{(p-1)p} = 0$$
$$4p - 3 = 0$$
$$\ddot{p} = \frac{3}{4}$$

We can observe that $\ddot{p}$ refers to the number of H in our 4 initial flips. In order to verify $\ddot{p}$ being a maximum, we take a look at the second derivative of $\log L(p)$ and look at its sign

$$\frac{d^2 \log L(p)}{d^2 p} = -\frac{4p^2-6p+3}{(p-1)^2 p^2} \leq 0$$

As the function is concave, it is clear that the value found is indeed a maximum.

e. Now for a little bit of practice manipulating conditional probabilities. Suppose that A and B are two events such that $P(A|B) = P(B|A)$. We also know that $P(A \cup B) = \frac{1}{2}$ and $P(A \cap B) > 0$. Prove that $P(A) > \frac{1}{4}$.

[**HINT:** Note that A and B are not necessarily mutually exclusive. Consider how we can relate $P(A \cup B)$ and $P(A \cap B)$.]

[**What we expect:** A short ($\sim$ 5 line) proof/derivation.]

Your Solution: Let's start by observing that $P(A|B) = P(B|A)$ means that $\frac{P(A \cap B)}{P(A)} = \frac{P(A \cap B)}{P(B)}$ and since $P(A \cap B) > 0$, we know that $P(A) = P(B)$. Additionally, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ from definition and, because $P(A) = P(B)$, then $P(A \cup B) = 2P(A) - P(A \cap B)$. Now we are finally able to correlate $P(A \cap B)$ and $P(A \cup B)$ with the following expression

$$\tfrac{1}{2} = P(A \cup B) > 2P(A) - P(A \cap B) \geq 2P(A)$$

and getting $P(A) > \frac{1}{4}$.

f. Let's practice computing gradients, which is a key operation for being able to optimize continuous functions. For $\mathbf{w} \in \mathbb{R}^d$ (represented as a column vector), and constants $\mathbf{a}_i, \mathbf{b}_j \in \mathbb{R}^d$ (also represented as column vectors), $\lambda \in \mathbb{R}$, and a positive integer n, define the scalar-valued function

$$f(\mathbf{w}) = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathbf{a}_i^\top \mathbf{w} - \mathbf{b}_j^\top \mathbf{w})^2\right) + \frac{\lambda}{2}\|\mathbf{w}\|_2^2,$$

where the vector is $\mathbf{w} = (w_1, \dots, w_d)^\top$ and $\|\mathbf{w}\|_2 = \sqrt{\sum_{k=1}^d w_k^2} = \sqrt{\mathbf{w}^T\mathbf{w}}$ is known as the L_2 norm. Compute the gradient $\nabla f(\mathbf{w})$.
[**RECALL:** The gradient is a d-dimensional vector of the partial derivatives with respect to each w_i:

$$\nabla f(\mathbf{w}) = \left(\frac{\partial f(\mathbf{w})}{\partial w_1}, \dots, \frac{\partial f(\mathbf{w})}{\partial w_d}\right)^\top.$$

If you're not comfortable with vector calculus, first warm up by working out this problem using scalars in place of vectors and derivatives in place of gradients. Not everything for scalars goes through for vectors, but the two should at least be consistent with each other (when $d = 1$). Do not write out summations over dimensions, because that gets tedious.]

[**What we expect:** An expression for the gradient and the work used to derive it. ($\sim$ 5 lines). No need to expand out terms unnecessarily; try to write the final answer compactly.]

Your Solution: Let's first consider two distinctive terms from the expression and differentiate them separately (since they two addends). In (1) we can observe that it is possible to apply the chain rule

$$(1)\tfrac{\partial}{\partial w_k}\left(\textstyle\sum_{i=1}^n\sum_{j=1}^n(\mathbf{a}_i^\top \mathbf{w} - \mathbf{b}_j^\top \mathbf{w})^2\right) = \textstyle\sum_{i=1}^n\sum_{j=1}^n 2(\mathbf{a}_i^T w - \mathbf{b}_j^T w)(a_i - b_j)$$

While (2) is more immediate

$$(2)\tfrac{\partial}{\partial w_k}\tfrac{\lambda}{2}\|\mathbf{w}\|_2^2 = \lambda w_k$$

Finally, the k^{th} element of $\nabla f(\mathbf{w})$ is

$$\tfrac{\partial f(\mathbf{w})}{\partial w_k} = \textstyle\sum_{i=1}^n\sum_{j=1}^n 2(\mathbf{a}_i^T w - \mathbf{b}_j^T w)(a_i - b_j) + \lambda w_k$$

Now, we can successfully express the gradient as follows

$$\nabla f(\mathbf{w}) = \left(\tfrac{\partial f(\mathbf{w})}{\partial w_1}, ..., \tfrac{\partial f(\mathbf{w})}{\partial w_d}\right) =$$
$$\left(\textstyle\sum_{i=1}^n\sum_{j=1}^n 2(\mathbf{a}_i^T w - \mathbf{b}_j^T w)(a_i - b_j) + \lambda w_1, ..., \sum_{i=1}^n\sum_{j=1}^n 2(\mathbf{a}_i^T w - \mathbf{b}_j^T w)(a_i - b_j) + \lambda w_d\right)$$

Problem 2: Complexity

When designing algorithms, it's useful to be able to do quick back-of-the-envelope calculations to see how much time or space an algorithm needs. Hopefully, you'll start to get more intuition for this by being exposed to different types of problems.

a. Suppose we have an $n \times n$ grid of points, where we'd like to place 5 arbitrary axis-aligned rectangles (i.e., the sides of the rectangle are parallel to the axes). Each corner of each rectangle must be one of the points in the grid, but otherwise there are no constraints on the location or size of the rectangles. For example, it is possible for all four corners of a single rectangle to be the same point (resulting in a rectangle of size 0) or for all 5 rectangles to be on top of each other. How many possible ways are there to place 5 rectangles on the grid? In general, we only care about asymptotic complexity, so give your answer in the form of $O(n^c)$ or $O(c^n)$ for some integer c.

[**NOTE:** It is unnecessary to consider whether order matters in this problem, since we are asking for asymptotic complexity. You are free to assume either in your solution, as it doesn't change the final answer.]

[**What we expect:** A big-O bound for the number of possible ways to place 3 rectangles and some simple explanation/reasoning for the answer ($\sim$ 2 sentences).]

Your Solution: We can start by considering the grid dimension $n \times n$ and the set of points (2) that defines the position of each rectangle, obtaining an initial $\mathbf{O}(n^4)$. There are 5 rectangles in total, and by scaling the previous result's exponent by that number, we get $\mathbf{O}(n^{20})$

b. Suppose we have an $n \times 2n$ grid of points. We start at the point in the upper-left corner (the point at position $(1,1)$), and we would like to reach the point at the lower-right corner (the point at position $(n,2n)$) by taking single steps down or to the right. Suppose we are provided with a function $c(i,j)$ that outputs the cost associated with position (i,j), and assume it takes constant time to compute for each position. Note that $c(i,j)$ can be negative. Define the cost of a path as the sum of $c(i,j)$ for all points (i,j) along the path, including both endpoints. Give an algorithm for computing the cost of the minimum-cost path from $(1,1)$ to $(n,2n)$ in the most efficient way (with the smallest big-O time complexity). What is the runtime (just give the big-O)?

[**What we expect:** A description of the algorithm for computing the cost of the minimum-cost path as efficiently as possible ($\sim$ 5 sentences). The big-O runtime and a short explanation of how it arises from the algorithm.]

Your Solution: First, let's define a matrix (2D array) M of $nx2n$ dimension. The $m_{ij} \in M$ element represents the smallest cost associated to get to the cell. The cost of the starting point (base case) equals to the cost of the cell itself

$$m_{11} = c(1,1)$$

Then, we can safely say that $c(i,j)$ to reach the ij^{th} cell follows the recursive pattern below

$$m_{ij} = c(i,j) + min(m_{(i-1)j}, m_{i(j-1)})$$

which will output the final m_{n2n} minimum cost. The time complexity $\mathbf{O}(n^2)$ follows the dimension of M $(2n^2)$

Problem 3: Ethical Issue Spotting

One of the goals of this course is to teach you how to tackle real-world problems with tools from AI. But real-world problems have real-world consequences. Along with technical skills, an important skill every practitioner of AI needs to develop is an awareness of the ethical issues associated with AI. The purpose of this exercise is to practice spotting potential ethical concerns in applications of AI - even seemingly innocuous ones.

In this question, you will explore the ethics of four different real-world scenarios using the ethics guidelines produced by a machine learning research venue, the NeurIPS conference. The NeurIPS Ethical Guidelines list fifteen non-exhaustive concerns under Potential Harms Caused by the Research Process and Societal Impact and Potential Harmful Consequences (containing the bulleted lists). For each scenario, you will write a potential negative impacts statement. To do so, you will first determine if the algorithm / dataset / technique could have a potential negative social impact or violate general ethical conduct (again, the fifteen items taken from the NeurIPS Ethical Guidelines page). If the scenario does violate ethical conduct or has potential negative social impacts, list one concern it violates and justify why you think that concern applies to the scenario. If you do **not** think the scenario has an ethical concern, explain how you came to that decision. Unlike earlier problems in the homework there are many possible good answers. If you can justify your answer, then you should feel confident that you have answered the question well.

Each of the scenarios is drawn from a real AI research paper. The ethics of AI research closely mirror the potential real-world consequences of deploying AI, and the lessons you'll draw from this exercise will certainly be applicable to deploying AI at scale. As a note, you are **not** required to read the original papers, but we have linked to them in case they might be useful. Furthermore, you are welcome to respond to anything in the linked article that's not mentioned in the written scenario, but the scenarios as described here should provide enough detail to find at least one concern.

[**What we expect:** A 2-5 sentence paragraph for each of the scenarios where you either A. identify at least one ethical concern from the NeurIPS Ethical Guidelines and justify why you think it applies, or B. state that you don't think a concern exists and justify why that's the case. Chosen scenarios may have anywhere from zero to multiple concerns that match, but you are only required to pick one concern (if it exists) and justify your decision accordingly. Furthermore, copy out and underline the appropriate parts of the ethical checklist item to which you are referring as part of your answer (e.g., promoting fossil fuel extraction from Societal Impact and Potential Harmful Consequences: Environment). We have also included a citation in the example solution below, but you are not required to add citations to your response.]

Example Scenario

You work for a U.S. hospital that has recently implemented a new intervention program that enrolls at-risk patients in programs to help address their chronic medical issues proactively before the patients end up in the hospital. The intervention program automatically identifies at-risk patients by predicting patients' risk scores, which are measured in terms of healthcare costs. However, you notice that for a given risk score tier, the Black patients are considerably sicker when enrolled than white patients, even though their assigned illness risk score is identical. You manually re-assign patients' risk scores based on their current symptoms and notice that the percentage of Black patients who would be enrolled has increased from 17% to over 45% [1].

Example Solution

This algorithm has likely encode, contain or exacerbate bias against people of a certain gender, race, sexuality, or other prote since the algorithm predicts healthcare costs. Because access to medical care in the U.S. is unequal, Black patients tend to have lower healthcare costs than their white counterparts [2]. Thus the algorithm will incorrectly predict that they are at lower risk.

a. An investment firm develops a simple machine learning model to predict whether an individual is likely to default on a loan from a variety of factors, including location, age, credit score, and public record. After looking through their results, you find that the model predicts mainly based on location and that the model mainly accepts loans from urban centers and denies loans from rural applicants [3]. Furthermore, looking at the gender and ethnicity of the applicants, you find that the model has a significantly higher false positive rate for Black and male applicants than for other groups. In a false positive prediction, a model misclassifies someone who does not default as likely to default.

Your Solution: The presented machine learning model has developed a clear aversion against people of certain gender and ethnicity, leading to a clear violation the bias and fairness. Furthermore the model erroneously distincts people from their ubication and will discriminate people by denying them credit access erroneously.

b. Stylometry is a way of predicting the author of contested or anonymous text by analyzing the writing patterns in the anonymous text and other texts written by the potential authors. Recently, highly accurate machine learning algorithms have been developed for this task. While these models are typically used to analyze historical documents and literature, they could be used for deanonymizing a wide range of texts, including code [4].

Your Solution: The mentioned model violates the privacy of the author responsible of intended anonymous text, one of the pillars in terms of human rights. This could potentially lead to the author's discrimination, social exclusion and even harassment.

c. A research group scraped millions of faces of celebrities off of Google images to develop facial recognition technology [5]. The celebrities did not give permission for their images to be used in the dataset and many of the images are copyrighted. For copyrighted photos, the dataset provides URL links to the original image along with bounding boxes for the face.

Your Solution: The above-mentioned model, represents again a perfect example of privacy violation, since the data collected for a data set, should always be authorized by its source, especially if the source is directly a human being. Providing an URL is not enough, the model should refers to uncopyrighted or at least authorized content.

d. Researchers have recently created a machine learning model that can predict plant species automatically directly from a single photo [6]. The model was trained using photos uploaded to the iNaturalist app by users who consented to use of their photos for research purposes, and the model is only used within the app to help users identify plants they might come across in the wild.

Your Solution: This ML model doesn't violate any of the ethical points. As long as the data set is documented, well-structured and non-misleading, it encourages curiosity and help humans in a innocuous way.

Problem 4: Programming

In this problem, you will implement a bunch of short functions. The main purpose of this exercise is to familiarize yourself with Python, but as a bonus, the functions that you will implement will come in handy in subsequent homeworks.

Do not import any outside libraries (e.g. numpy). Only standard python libraries and/or the libraries imported in the starter code are allowed.

See `submission.py`. No written submission.

Submission

Submission is done on Gradescope.

Written: When submitting the written parts, make sure to select **all** the pages that contain part of your answer for that problem, or else you will not get credit. To double check after submission, you can click on each problem link on the right side and it should show the pages that are selected for that problem.

Programming: After you submit, the autograder will take a few minutes to run. Check back after it runs to make sure that your submission succeeded. If your autograder crashes, you will receive a 0 on the programming part of the assignment. Note: the only file to be submitted to Gradescope is `submission.py`.

More details can be found in the Submission section on the course website.

References

[1] Obermeyer et al. Dissecting racial bias in an algorithm used to manage the health of populations. 2019.

[2] Institue of Medicine of the National Academies. Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care. 2003.

[3] Imperial College London. Loan Default Prediction Dataset. 2014.

[4] Caliskan-Islam et. al. De-anonymizing programmers via code stylometry. 2015.

[5] Parkhi et al. VGG Face Dataset. 2015.

[6] iNaturalist. A new vision model. 2020.