

CS221 Spring 2025: Artificial Intelligence: Principles and Techniques

Homework 1: Foundations

SUNet ID: [Kiranmm]
Name: [Kiran Makhijani]
Collaborators: [-]

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 Homeworks section on the course website thoroughly.

Prerequisite Quiz

Please take the prerequisite quiz found on Gradescope. The quiz has the same due date as HW1. The quiz is open book, you have unlimited attempts, and there is no time limit. However, you will still be given a grade for it based on correctness.

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. The textbook for MATH 51 may be useful for the optimization problems here, specifically the sections “Maxima, Minima, and Critical Points” (p. 186) and “Gradients, Local Approximations, and Gradient Descent” (p. 209).

At the same time, we must be adroit at coping with uncertainty in the world, and for that, we appeal to tools from probability. The course reader for CS109 may be useful for the probability problems here, specifically the sections “Conditional Probability”, “Expectation”, “Joint Probability”, and “Bayes’ Theorem”.

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:

$$\begin{aligned} f(\theta) &= \sum_{i=1}^n w_i(\theta - x_i)^2 \\ &= \sum_{i=1}^n w_i(\theta^2 + x_i^2 - 2\theta x_i) \end{aligned}$$

first we find critical points by taking derivative and setting it to 0.

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

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

Since w_i are positive, above equation will be minimum for 0 when $\theta = x_i$

$$\text{Or: } \theta = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

Intuitively, if w_i 's were negative then sum could also be negative and the relationship between scalar and x_i can not be determined.

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 $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ 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 rewrite the functions so that they are minimizing the same quantity. The way to do this would involve defining the coefficients of s_i for all terms in the summation. Then examine how the constraints on s_i for $f(\mathbf{x})$ differ from those for $g(\mathbf{x})$]

[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 proof may be somewhat informal, but please describe your steps and thought process clearly.]

Your Solution:

We start with expanding the function $f(\mathbf{x})$ for values of s .

$f(\mathbf{x}) = \min_{s \in [-1, 1]} \sum_{i=1}^d s x_i$ is same as:

$$f(\mathbf{x}) = \min(-1 \cdot \sum_{i=1}^d x_i, 0, 1 \cdot \sum_{i=1}^d x_i)$$

Now sum of the x_i for $x = 1 \dots d$ can be positive or negative. But the absolute value of both expressions with -1 and $+1$ is positive. We also note that since there is a negative value, 0 is not the minimum and we can ignore this. In other words, we can say $f(\mathbf{x}) = -\left|\sum_{i=1}^d x_i\right|$. The same way we expand $g(\mathbf{x})$ to regroup the expression as follows

$$g(\mathbf{x}) = \sum_{i=1}^d \min_{s_i \in [-1, 1]} s_i x_i$$

$$g(\mathbf{x}) = \sum_{i=1}^d \min(-1 \cdot x_i, 0, 1 \cdot x_i)$$

$$g(\mathbf{x}) = \sum_{i=1}^d -1 \cdot |x_i|$$

$$g(\mathbf{x}) = -1 \cdot \sum_{i=1}^d |x_i|$$

$$\text{Let } S_i = \sum_{i=1}^d |x_i| \text{ and } S_o = \left|\sum_{i=1}^d x_i\right|$$

To answer the question, we compare S_o and S_i

- $f(\mathbf{x}) = g(\mathbf{x})$: This is True when $S_o = S_i$
- $f(\mathbf{x}) \leq g(\mathbf{x})$: True when $S_o \leq S_i$
- $f(\mathbf{x}) \geq g(\mathbf{x})$: True when $S_o \geq S_i$

Intuitively, an example is that when all x_i have the same sign (+/-), then they are equal. Otherwise, the answer depends on the sum of x_i s. For example, for $\{3, 4, -5\}$ S_o is 2 and S_i is 12. This shows impact of order of operations.

Suppose you repeatedly roll a fair six-sided die until you roll a 1 (and then you stop). Every time you roll a 3, you win a points, and every time you roll a 6, you lose b points. You do not win or lose any points if you roll a 2, 4 or a 5. 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? More precisely, can you write an equation that relates V to an expression involving itself, which can then be solved to find the value of V ?]

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

Your Solution: Here, we compute probability of each event and associated value with it. since score 1 is terminating case, it has no value, at 2,4,5 there is no change in value so

$V = P_1 * 0 + P_{245} * V + P_3(V + a) + P_6(V - a)$ then

$V = \frac{1}{6} * 0 + \frac{3}{6}(V) + \frac{1}{6}(V + a) + \frac{1}{6}(V - b)$

$6V = 3V + V + a + V - b$

$6V - 5V = a - b$ or

$V = a - b$

Suppose the probability of a coin turning up heads is p (where $0 < p < 1$), and we flip it 5 times and get $\{T, H, H, H, H\}$. We know the probability (likelihood) of obtaining this sequence is $L(p) = (1 - p)pppp = p^4(1 - p)$. What value of p maximizes $L(p)$? Prove 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 that it is a maximum. A 1-sentence intuitive interpretation of the value of p .]

Your Solution:

Lets take derivative of $\log L(p)$

$$L(p) = (1 - p)pppp$$

$$\log(L(p)) = \log(1 - p) + \log(p^4)$$

$$\frac{\partial \log(L(p))}{\partial p} = -\frac{1}{1-p} + 4\frac{1}{p}$$

solving for derivative set to 0 and get:

$$-\frac{1}{1-p} + 4\frac{1}{p} = 0$$

$$-p + 4(1 - p) = 0$$

$$\text{Or } p = \frac{4}{5}$$

from this equation there's one solution. to further check maximum, perform the second derivative test.

$$\frac{\partial}{\partial p} \frac{1}{1-p} = \frac{1}{(1-p)^2}$$

and

$$\frac{\partial}{\partial p} \frac{1}{p} = -\frac{1}{p^2}$$

second derivative is

$$= -\frac{1}{(1-p)^2} - \frac{1}{p^2}$$

since $p > 0$, the second derivative test is < 0 . So it must be maxima.

Intuitively, it seems obvious that in a given sequence there were 5 flips, 4 times it gave H which is $4/5$ (likelihood of H over total number of flips). Max probability of 1 is not possible as one time H did not occur.

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 $P(A \cup B) = \frac{1}{3}$ and $P(A \cap B) > 0$. Prove that $P(A) > \frac{1}{6}$. [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 (~ 5 line) proof/derivation.]

Your Solution: Given:

$$P(A|B) = P(B|A) \text{ and } P(A \cup B) = \frac{1}{3}$$

We use 2 formulas.

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

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

this means $P(A) = P(B)$ and using this result in first

$$P(A \cup B) + P(A \cap B) = 2P(A)$$

$$\frac{1}{3} + P(A \cap B) = 2P(A)$$

$$\frac{1}{6} + P(A \cap B)\frac{1}{2} = P(A)$$

Since we know that $P(A \cap B) > 0$, it shows that $P(A)$ is $1/6$ and some non-zero positive value.

$$P(A) > \frac{1}{6}$$

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 $\|\mathbf{w}\|_2 = \sqrt{\sum_{k=1}^d w_k^2} = \sqrt{\mathbf{w}^\top \mathbf{w}}$ 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. (~ 5 lines). No need to expand out terms unnecessarily; try to write the final answer compactly.]

Your Solution:

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.

Suppose we have an $n \times n$ grid of points, where we'd like to place 3 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 3 rectangles to be on top of each other. How many possible ways are there to place 3 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 (~ 2 sentences).]

Your Solution: The number of ways 3 points will be on top of each other = $O((n-2)*n)$ or $O(n^2)$. There can be n different sizes of rectangles, and we can stack each set of 3 rectangle in roughly $(n-2)$ ways. Thus total possible ways are: $O(n^2) + O(n^2)$ which is $O(n^2)$.

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 (~ 5 sentences). The big-O runtime and a short explanation of how it arises from the algorithm.]

Your Solution: Bellman Ford is a well known graph algorithm to compute least-cost path. We will first convert the grid to a graph of vertices and edges. Each point represents a vertex in the graph and there will be two outgoing edges to the right and bottom. for vertex (i, j) , edges are $(i + 1, j)$ and $(i, j + 1)$. Once the $V = 2n * n$ and $E = 2V$.

In Bellman ford, cost-distance to each vertex is set to ∞ initially and as edges are traversed cost-distance to vertex is updated using $\text{dist}[j] = \min(\text{dist}[j], \text{dist}[i] + w)$. The basic idea is that each edge must be traversed; therefore, this step is repeated $V - 1$ times.

This is not a greedy algorithm, and is known to be quite efficient. The runtime complexity of the algo is $O(V * V - 1)$ ($V-1$ represents maximum outgoing edges) or in this case $O(n^2 * n^2)$ or $O(n^4)$.

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 seventeen non-exhaustive concerns under General Ethical Conduct and Potential Negative Social Impacts (the numbered 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 seventeen numbered 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; you should think about why the researchers may have chosen for the algorithms to behave in the way as described in the scenario. 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 ethical checklist item to which you are referring as part of your answer (i.e.: Severely damage the 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 encoded, contains, or potentially exacerbates bias against people of a certain race or ethnicity 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: This algorithm also has likely encoded, contains, or potentially exacerbates bias against people of a certain race. Because the dataset does not fairly represent the diversity of the rural community.

The model predicts ability to pay loan. High false positives are seen because the credit score is more relevant in urban areas than rural areas, rural workers will have lower credit scores. The dataset has high false positive for Black families (or individual) because it does not take into account that mean income among Black families is lower. Lower income does not relate to default the loans. The model is biased to location mainly, i.e. accepts loans from urban and denies loans to rural community shows that it has not factored in income gap between rural and urban regions. Farmers or workers needing loans in rural areas will have lower income. Since most individuals seeking loans are males, the dataset does not seem to balance the gap between number of male applicants vs female applicant. Thus the algorithm will incorrectly predict that women are less likely to default.

- 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 concern is likely related to Have a detrimental effect on people's livelihood or economic security. As the abstract of [4] starts off with 'Source code authorship attribution is a significant privacy threat to anonymous code contributors'. An author or code writer may have valid reasons to preserve their privacy from their corporations or state. It is possible that such predictions can lead to disclosing their identity which in extreme cases may cause harm to them.

- 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 likely concern here is Contain information that could be deduced about individuals that they have not consented to share. Since the images are copyrighted, a pointer to image URL does not imply that celebrities have consented to share [7].

- 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: I do not have strong concern in building dataset to predict plant species. There is no obvious harm to this model and the core data information is about the plants. Although there maybe privacy concerns which could expose person's identifiable information such as location (geo coordinates) when uploading surrounding area coordinates.

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] Institute 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.
- [7] VGG Face dataset used personal data without explicit consent from the individuals depicted.