CS221 – Artificial Intelligence: Principles and Techniques

Homework 1: Foundations

Autumn 2025

SUNet ID: arzyn
Name: Artem Arzyn
Collaborators: Min Sik (OH)

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

Before you get started, please read the Homeworks section on the course website thoroughly. This LaTeXtemplate serves as a convenient template for the written parts of the assignment. The questions mirror those in the HTML page; if there are any typos, follow the HTML and let us know in your answers.

Problem 1: Linear Algebra

Linear algebra forms the foundation of modern AI and machine learning. In this problem, you'll work with vectors and matrices, learn basic NumPy and einsum, and write compact expressions for common operations.

a. **Learn basic NumPy operations with an AI tutor!** Use an AI chatbot (e.g., ChatGPT, Claude, Gemini, or the Stanford AI Playground) to teach yourself basic vector and matrix operations in NumPy (`import numpy as np`). If you'd like more help, this YouTube video may be helpful.

[**What we expect:** Provide a link to the chat session transcript with the AI tutor. The session should be 15–20 minutes and interactive.]

Your Solution: Link:
https://drive.google.com/file/d/1Z81rcCVk8jMhlV8DbVXb51SbxEExVjWy/view?usp=drive$_l ink$

b. **Linear Algebra Complexity.** Suppose you have two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times p}$. What is the time complexity of computing their product AB using the standard matrix multiplication algorithm? Express your answer using big-O notation and briefly justify why.

[**What we expect:** The time complexity in big-O notation and a 1–2 sentence explanation.]

Your Solution: O(m*n*p) since the complexity is based on multiplying the ith row of Matrix A by the jth column of matrix B in order to calculate ij element of the product AB. Effectively, we need to calculate m*p number of elements from matrix AB which requires O(n) operations each to compute based on the mxn and nxp dimensions.

c. **Learn basic einsum with an AI tutor!** Use an AI chatbot to teach yourself Einstein summation notation (einsum) in NumPy.

[**What we expect:** Provide a link to the chat session transcript with the AI tutor. The session should be 15–20 minutes and interactive.]

Your Solution: Link:
https://drive.google.com/file/d/1VmfOL-llW1EDkLnRXWGQqVf7BxJmz22A/view?usp=drive$_l ink$

d. **Einstein Summation (Written).** Given $X \in \mathbb{R}^{n\times d}$ and $\mathbf{w} \in \mathbb{R}^d$: (i) write an `einsum` string for $X\mathbf{w}$; (ii) for the pairwise dot-product matrix $XX^\top$; (iii) for $\text{diag}(X^\top X)$ (column-wise squared norms). Briefly justify each.

[**What we expect:** Provide einsum strings (e.g., `n d, d m -> n m`).]

Your Solution:

i. `'n d, d -> n'`, based on X of shape (n, d) multiplied by w of shape (d,), so we end with a shape (n,) with rows dot producted with the w vector.

ii. `'n d, m d -> n m'` since X has shape (n, d) and X^T has shape (m, d) where m = n numerically. The X^T has (m, d) indices since we have rows which become columns post transpose. We have the rows of X (the n d one) and dot product from X^T with the d shared dimension, so the final result is of shape (n, m) which is the same size as (n, n).

iii. `'n d, n d -> d'` Since we want diagnals from $X^T X$, which would involve using the same indices, we then get the diagonal elements for sum of squares per column.

Problem 2: Calculus and Gradients

Gradients are essential for training machine learning models through optimization algorithms like gradient descent. In this problem, you'll practice computing gradients analytically.

a. **Gradient Warmup.** For $f(\mathbf{w}) = \sum_{i=1}^{d}(w_i - c_i)^2$, derive $\nabla f(\mathbf{w})$. Then evaluate the gradient at $\mathbf{w} = \mathbf{0}$.

[**What we expect:** A compact vector expression (e.g., $(\mathbf{w} + \mathbf{c})$) and one evaluated vector.]

Your Solution: We are given $f(\mathbf{w}) = \sum_{i=1}^{d}(w_i - c_i)^2$. We must find $\nabla f(\mathbf{w})$ and evaluate the gradient at $\mathbf{w} = \mathbf{0}$.

To find $\nabla f(\mathbf{w})$, we can compute the partial derivative with respect to each component. This means, we find the derivative with respect to each w_j. First we have $\frac{\partial}{\partial w_j}(w_j - c_j)^2 = 2(w_j - c_j)$. Then since $\sum_{i=1}^{d}(w_i - c_i)^2$ has terms which don't depend on w_j, we remove those when doing this partial derivative with respect to w_j. This means only the j-th term $(w_j - c_j)^2$ remains.

Thus,

$$\frac{\partial f}{\partial w_j} = 2(w_j - c_j)$$

which means $\nabla f(\mathbf{w}) = 2(\mathbf{w} - \mathbf{c})$

At $\mathbf{w} = \mathbf{0}$, this is $\nabla f(\mathbf{w}) = 2(\mathbf{w} - \mathbf{c}) = 2(\mathbf{0} - \mathbf{c}) = -2\mathbf{c}$.

c. **Matrix Multiplication Gradient** Consider two matrices A (size $m \times n$) and B (size $n \times p$) that are multiplied together to form $C = AB$, and then all entries of C are summed to produce a scalar $s = \sum_{i,j} C_{i,j}$.

Let $A = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 8 \\ 9 & 0 \\ 1 & 2 \end{pmatrix}$.

Compute $C = AB$ and $s = \sum_{i,j} C_{i,j}$. Then find the gradient $\frac{\partial s}{\partial A_{i,k}}$ for each entry of matrix A, and similarly find $\frac{\partial s}{\partial B_{k,j}}$ for each entry of matrix B.

[**What we expect:** The computed matrices C and scalar s, plus the gradient matrices $\frac{\partial s}{\partial A}$ and $\frac{\partial s}{\partial B}$ with numerical values for each entry.]

Your Solution: We are provided $A = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 8 \\ 9 & 0 \\ 1 & 2 \end{pmatrix}$. We must find $C = AB$, $s = \sum_{i,j} C_{i,j}$, $\frac{\partial s}{\partial A_{i,k}}$ for each entry of matrix A, and find $\frac{\partial s}{\partial B_{k,j}}$ for each entry of matrix B.

i. $C = AB$
Due to the dimensions of A and B, we know C is 2×2 in size. $C_{11} = A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31} = 2 * 7 + 1 * 9 + 3 * 1 = 14 + 9 + 3 = 26$. By the same process, we find $C_{12} = 22$, $C_{21} = 79$, and $C_{22} = 44$ which makes
$$C = \begin{pmatrix} 26 & 22 \\ 79 & 44 \end{pmatrix}$$

ii. $s = \sum_{i,j} C_{i,j}$
Based on C above, $s = 26 + 22 + 79 + 44 = 171$

iii. $\frac{\partial s}{\partial A_{i,k}}$ for each entry of matrix A To find $\frac{\partial s}{\partial A_{i,k}}$ based on entries of matrix A, we have to take the partial derivative with respect to $A_{i,k}$. $\frac{\partial s}{\partial A_{i,k}} = \frac{\partial}{\partial A_{i,k}} \sum x, y \sum z A_{x,z} B_{z,y} = \sum_j B_{k,j}$. Here, we see the entries of $\frac{\partial s}{\partial A_{i,k}}$ are equal to the sums of k-th rows of matrix B. Specifically, this means the first entry would be $B_{11} + B_{12} = 7 + 8 = 15$ and onwards for the other rows. $\frac{\partial s}{\partial A_{i,k}} = \begin{pmatrix} 15 & 9 & 3 \\ 15 & 9 & 3 \end{pmatrix}$

iv. $\frac{\partial s}{\partial B_{k,j}}$ for each entry of matrix B To find $\frac{\partial s}{\partial B_{k,j}}$ based on entries of matrix B, we have to take the partial derivative with respect to $B_{k,j}$. $\frac{\partial s}{\partial B_{k,j}} = \frac{\partial}{\partial B_{k,j}} \sum x, y \sum z A_{x,z} B_{z,y} = \sum_j B_{k,j}$. Here, we see the entries of $\frac{\partial s}{\partial B_{k,j}}$ are equal to the sums of k-th columns of matrix A. Specifically, this means the first entry would be $A_{11} + A_{21} = 2 + 4 = 6$ and onwards for the other columns. $\frac{\partial s}{\partial B_{k,j}} = \begin{pmatrix} 6 & 6 \\ 6 & 6 \\ 9 & 9 \end{pmatrix}$

Problem 3: Optimization

Optimization is central to AI. In this problem, you'll work with analytical optimization and relate it to numerical methods.

a. Let $x_1, \ldots, x_n$ be real numbers and $w_1, \ldots, w_n$ positive weights. Consider $f(\theta) = \sum_{i=1}^{n} w_i(\theta - x_i)^2$ (scalar θ). What value of θ minimizes $f(\theta)$? Show it is a minimum. What issues arise if some w_i are negative?

[**What we expect:** Closed-form minimizer, brief justification it's a minimum, and a note on negative weights.]

Your Solution: We must find the value of θ that minimizes $f(\theta)$ and show it is a minimum, as well as describe issues that arise if some w_i are negative.

To find the minimum, we take the derivate with respect to θ and set it equal to zero. $\frac{\partial f}{\partial \theta} = \sum i = 1^n w_i \cdot 2(\theta - x_i) = 2\sum_{i=1}^n w_i(\theta - x_i) = 2\theta\sum_{i=1}^n w_i - 2\sum_{i=1}^n w_i x_i$

From there, $2\theta\sum_{i=1}^n w_i - 2\sum_{i=1}^n w_i x_i = 0$, solving for θ gets us $\theta^{opt} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$, the weighted average of positions x_i with weights w_i. To check for it being a minimum, we then take the second derivative and must confirm it is greater than zero. $\frac{\partial^2 f}{\partial \theta^2} = 2\sum i = 1^n w_i$. Since all $w_i > 0$, we know $2\sum i = 1^n w_i > 0$ and, therefore, that it is a minimum. If some w_i were negative, the function might not necessarily have a negative since $2\sum i = 1^n w_i \leq 0$ would then be a possibility, meaning a non-positive second derivative and, therefore, not a minimum.

b. **Learn about gradient descent with an AI tutor!** Use an AI chatbot to teach yourself gradient descent optimization techniques.

[**What we expect:** Provide a link to the chat session transcript with the AI tutor. The session should be 15–20 minutes and interactive.]

Your Solution: Link: https://drive.google.com/file/d/1Ik-a3980ptEWeXV5-6mX-EK214s6Irmi/view?usp=drive$_l ink$

Problem 4: Ethics in AI

For this problem, write brief responses for four ethics scenarios. Please refer to the assignment page (index.html) for the full scenario text and detailed expectations.

a. Scenario 4a (refer to index.html for the description).

Your Solution: This scenario specifically raises human rights concerns and could have a detrimental effect on people's livelihood or economic security. As we have seen before, removing information on gender and ethnicity does not remove other approximations of identity (previous ethical questions about artificial intelligence have come up for exactly this reason where zip code and income can correlate enough with gender and ethnicity that removal of the latter two to avoid discrimination is ineffective). This algorithm could easily have been trained on a dataset that is resulting in the exacerbation of bias against certain groups, even while excluding that information, and in the case of rural applicants, the inclusion of information could be contributing to bias. This could be used to discriminate against Black, male, or rural applicants, and could also exacerbate existing biases since denial of loans or classification of someone as more likely to default than they are could result in them having to turn to more predatory lending services, pay higher interest rates, or go without funds at all.

b. Scenario 4b (refer to index.html for the description).

Your Solution: This algorithm develops or extends harmful forms of surveillance. Specifically, denanonymizing writing without consent could cause harm by deanonymizing personal content on anonymous social media platforms, anonymous open source code for activism uses, anonymous whistleblowing witnesses, and broadly just put people in a position to be attached to information they never expected to be attached to without their consent. This is a violation of privacy, could cause issues with safety, and could result in harassment of individuals who speak or type under condition of anonymity.

c. Scenario 4c (refer to index.html for the description).

Your Solution: This one specifically violates principles of consent and doesn't contain any restrictions on the use of the dataset, so could also develop or extend harmful forms of surveillance and bring up human rights concerns. Specifically, the fact that we are using photos of celebrities who are people without their consent, using photographs without photographers' consent (even if they include URLs for copyrighted images), and are building a facial recognition technology with no specific restrictions in place means that the project is violating norms about ethical research.

d. Scenario 4d (refer to index.html for the description).

Your Solution: In this scenario, they have obtained consent from participants, they have properly informed them of the use case of their data, and that use case is being met, so there do not appear to be ethical issues directly. There could be some hypothetical risks around identification of plants and people possibly assuming the model is 100% accurate in identification, such that they are harmed by consuming plants they ought not consume, or security risks with data collected including geolocated data, but as this scenario is presented, there do not appear to be ethical concerns

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.