Stanford CS 229, Winter 2024 Midterm

The midterm is open-book (no internet), closed-collaboration, and subject to the Honor Code. Please write your name and SID on the first page below and **write your name initials on the top of every page. Please take a look at the instructions below before you start.** There are **26 pages** in total. You **may**:

- You have an additional 15 minutes (totaling 3 hours and 15 minutes) to scan and upload your answers to Gradescope. **Do not use this time to work on the exam.**
- You can print and handwrite or write directly onto the exam using your digital device at your convenience. You can also write your answer by hand on blank pages, take photos, copy your photos to the exam booklet, and then upload the complete booklet to Gradscope. **In all cases, please make sure your answers are on the blank exam booklet instead of completely blank pages. Please upload the exact number of pages (including the last blank pages) given in the exam booklet.** You don't need to select pages.
- Access any materials or resources, including the course notes and reference material you may have downloaded or printed previously. You can use electronic devices, but **cannot connect to the internet**.
- Cite without proof any result from lecture slides, homework, or lecture notes, unless otherwise stated.
- If you encounter a question that needs clarification, please write your assumptions at the beginning of your solution for the teaching staff to consider when grading. For example, "I wasn't sure if this question was asking X or Y. I assumed X and answered the question accordingly."

You **may not**:

- Talk to, consult, or collaborate with anyone about the exam, and you may not consult any human or artificial intelligence about the exam problems. Any such collaboration is a violation of the Honor Code.
- **Access any resources from the internet during the midterm.** Note: you are allowed to download materials from the internet ahead of the midterm and access them offline during the midterm.
- You are not allowed to ask questions on Ed during the exam and we will not answer any clarification questions.

Good luck! We know you've been working hard, and we all want you to succeed!

Name of Student: ______________________________

SUNetID: ______________________________@stanford.edu

Exam Duration: 3 hours

Question	Points
1 True or False	/12
2 Multiple Choice	/12
3 Decision Tree Construction	/14
4 Binomial GLM	/15
5 Tikhonov Regression	/10
6 Quadratic Loss Descent	/14
7 Multi-task Networks	/23
Total	/100

The Stanford University Honor Code:

I attest that I have not given or received aid in this examination, and that I have done my share and taken an active part in seeing to it that others as well as myself uphold the spirit and letter of the Honor Code. In addition, I have not accessed any online resources during the exam.

Signed: ______________________________

1. **[12 points] True or False**

For each statement, just indicate whether it is TRUE (always completely correct) or FALSE (at least some aspect is sometimes wrong). Make sure to fully fill out the checkboxes, i.e. ■. Answers which do not fully fill out the checkboxes, i.e. ☑ or ☒ may be marked incorrect. **No need to provide an explanation.**

(a) [2 points] Consider the ordinary least squares solution θ^* in $X^TX\theta^* = X^Ty$, where $X \in \mathbb{R}^{m\times d}$ consists of m examples with d features each. So long as $m \gg d$, i.e. the number of examples is way more than the number of features, there exists a unique least squares solution.

☐ True

☐ False

(b) [2 points] Consider a classification problem with the provided training dataset:

x_1	x_2	y
-1	2	0
0	3	0
1	4	0
-1	5	1
0	6	1
1	7	1

We perform logistic regression on the dataset, using the hypothesis form $h_\theta(x) = g(\theta^T x) = \frac{1}{1+e^{-\theta^T x}}$, where $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$. Using this parameterization, we can achieve 100% accuracy on the training dataset.

☐ True

☐ False

(c) [2 points] Given a unique global optimum, gradient ascent applied to the log-likelihood function of any Generalized Linear Model (GLM) is guaranteed to converge to this global optimum, assuming the step size is sufficiently small.

☐ True

☐ False