PRACTICE FINAL EXAM - DSC 40A, Fall 2024

Full Name:

PID:

Seat Number:

Instructions:

- This exam consists of 7 questions, worth a total of one million points.
- **Advice: Read all of the questions before starting to work, because the questions are not sorted by difficulty.**
- Write your PID in the top right corner of each page in the space provided.
- Please write **clearly** in the provided answer boxes; we will not grade work that appears elsewhere.
  - For questions that ask you to show your work, correct answers with no work shown will receive no credit.
- You may use two pages two-sided as a cheat sheet. Other than that, you may not refer to any resources or technology during the exam (no phones, no smart watches, no computers, and no calculators).

By signing below, you are agreeing that you will behave honestly and fairly during and after this exam.

Signature:

Version A

Please do not open your exam until instructed to do so.

Question 1 Multiple Linear Regression

Zoe knows you are all stressed because finals week is coming up! She understands and has been doing a lot of stress baking due to her own finals.

Zoe has to focus on time management when baking, yet strangely enough, she found her baking time follows a multivariate linear equation.

Her time for baking depends on the following information:

- Stressed?: x_1
- Number of Group Meetings: x_2
- Number of Assignments Due: x_3

Zoe keeps a record of her baking habits over four days. Each day, she records if she is stressed, the number of meetings she has that day, and how many assignments are due that day. The time for baking (B) for each day is given below:

Day	x_1	x_2	x_3	B
Mon	1	1	2	-2
Tues	0	2	1	4
Weds	1	0	1	4
Thurs	0	1	1	3

Notice: When she is super busy she does not have time to bake and when she does anyways it means on Monday she had two less hours to do her work!

Zoe believes her baking time can be modeled as:

$$B = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_3$$

where w_0 is the intercept, and w_1, w_2, w_3 are the coefficients for the variables.

a) Construct the **design matrix** X for the given data.

b) Use the normal equation for multiple linear regression to compute the coefficient vector $\vec{w}$ where y is the vector of dessert outputs B.

Note:

$$(X^T X)^{-1} = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 6 & 3 & -4 \\ -1 & 3 & 2 & -2 \\ -1 & -4 & -2 & 4 \end{bmatrix}$$

c) Calculate the residuals for the original dataset and then write them into the Mean Squared Error equation. This means you do not need to do these calculations by hand, but need to write it out!

Question 2 Bayes' Theorem

Zoe has decided to run a bakery in her free time and wants to predict whether a customer will buy a cookie, cake, or bread based on their preferences. She has collected the following data:

- Sweetness Preference: Low, Medium, High
- Texture Preference: Soft, Crunchy
- Occasion: Casual, Celebration

Zoe was able to collect some data from her customers: Gal, Brighten, Javier, Rebecca, Utkarsh, Owen, and Varun.

Sweetness	Texture	Occasion	Baked Good
High	Crunchy	Casual	Cookie
Medium	Soft	Casual	Bread
High	Soft	Celebration	Cake
Low	Crunchy	Casual	Bread
High	Crunchy	Celebration	Cookie
Medium	Soft	Celebration	Cake
Low	Soft	Casual	Bread

a) What is the probability for each baked good? (i.e. What is the probability of baking a cookie, a cake, a bread?)

b) What kind of baked good will Sawyer order if he wants a baked good with **Medium Sweetness** and a **Soft Texture**? Do **not** use smoothing.

c) What kind of baked good will Masfiqur order if he wants a baked good with **Low Sweetness** for a **Celebration**? **Use** smoothing.

PID: ______________________________

Question 3 Combinatorics

Zoe's kitchen has 10 different ingredients to choose from: flour, sugar, butter, eggs, chocolate chips, vanilla, oatmeal, cinnamon, raisins, and pecans.

a) Zoe decides that her cookies must include at least one of the following "core" ingredients: flour, sugar, or butter. If she chooses 4 ingredients at random, what is the probability that her selection includes at least one of these core ingredients? Leave your answer unsimplified.

b) Suppose Zoe already picked one of her ingredients, which is chocolate chips (because Gal likes chocolate). What is the probability that the remaining 3 ingredients she picks will not include any of the core ingredients (flour, sugar, butter)? Leave your answer unsimplified.

c) What is the probability that Zoe's selection includes at least one sweet ingredient? Sweet ingredients are: sugar, chocolate chips, raisins, and vanilla. Leave your answer unsimplified.

d) If Zoe's selection is guaranteed to include exactly two sweet ingredients, what is the probability that the other two ingredients she selects are both grains (flour and oatmeal are considered grains)? Leave your answer unsimplified.

Question 4 Combinatorics

Zoe has baked 8 different desserts for her friends to enjoy: 2 pies, 2 cakes, 3 kinds of cookies, and 1 type of bread. She wants to display them on the table in a single row.

a) How many distinct arrangements of the baked goods are possible? You can leave the answers unsimplified.

b) If the 3 kinds of cookies must be next to each other (as if they were a single unit), how many distinct arrangements are possible? You can leave the answers unsimplified.

c) How many permutations of the arrangement contain all of the desserts that start with c before all of the others? (Cake and Cookies start with c).

Question 5 Multiple linear regression

DISCLAIMER This problem was originally going to be on your final, but we removed it because it was a bit too difficult. So don't panic if it seems hard; HOWEVER, it is an EXCELLENT review for understanding gradients, loss minimization, and multivariate regression -Sawyer

Consider the usual setup for a multiple regression problem, as follows. Let $\vec{x}1, \vec{x}_2, \ldots, \vec{x}_n \in \mathbb{R}^d$ be a collection of feature vectors in $\mathbb{R}^d$, and let $X \in \mathbb{R}^{n\times(d+1)}$ be the corresponding design matrix. Assume $(X^TX)^{-1}$ exists. Each $\vec{x}_i$ has a label $y_i \in \mathbb{R}$, and let $\vec{y} \in \mathbb{R}^n$ be the vector of labels. We wish to fit the data and labels with a linear model

$$H(\vec{x}) = w_0 + w_1\vec{x}^{(1)} + w_2\vec{x}^{(2)} + \ldots + w_d\vec{x}^{(d)} \ = \vec{w}^T \text{Aug}(\vec{x}).$$

where $\vec{w} \in \mathbb{R}^{d+1}$ is the vector of coefficients with intercept. Let $\beta_1, \beta_2, \ldots, \beta_n > 0$ be a list of strictly positive real numbers. Define the **weighted β-loss** as follows:

$$R_\beta(w) = \frac{1}{n}\sum_{i=1}^{n} \beta_i |y_i - H(\vec{x}_i)|^2.$$

Let $B \in \mathbb{R}^{n\times n}$ denote the $n \times n$ diagonal matrix of weights $\beta_1, \ldots, \beta_n$.

a) Prove that

$$R_\beta(w) = \frac{1}{n}\|B^{1/2}(\vec{y} - X\vec{w})\|^2,$$

where $B^{1/2} \in \mathbb{R}^{n\times n}$ is the diagonal matrix with entries $\beta_i^{1/2}$.

b) Let $A \in \mathbb{R}^{n\times(d+1)}$ be any matrix, and $\vec{a} \in \mathbb{R}^n$ be any fixed vector. For $\vec{z} \in \mathbb{R}^{d+1}$, define the function $F(\vec{z})$ by

$$F(\vec{z}) = \|\vec{a} - A\vec{z}\|^2,$$

Show that

$$\nabla_{\vec{z}} F(\vec{z}) = -2A^T(\vec{a} - A\vec{z})$$

Hint: First, show that

$$\frac{\partial}{\partial \vec{z}^{(j)}}(\vec{a} - A\vec{z})^{(i)} = -A_{ij},$$

and then find a formula for $\|\vec{a} - A\vec{z}\|^2$ *in terms of* $(\vec{a} - A\vec{z})^{(i)}$.

c) Use the previous part to prove that

$$\nabla R_\beta(\vec{w}) = -\frac{2}{n}\left(X^TB\vec{y} - X^TBX\vec{w}\right),$$

and find a formula for the optimal parameter vector $\vec{w}^*$ with respect to the loss R_β.

d) Imagine you are analyzing performance metrics from a toy factory to predict the proportion of broken toys y_i using several production factors $\vec{x}_i$ such as machine speed, worker experience, and material quality. Each data point represents a different batch of toys. Explain why it might make sense to incorporate weights β_i into the mean-squared error.

Question 6 Convexity

Assume $f(x)$ is a function which satisfies $f(x) > 0$ for all x. We say that $f(x)$ is **logarithmically convex** if $g(x) = \ln(f(x))$ is a convex function. Similarly, we say that $f(x)$ is **logarithmically concave** if $g(x) = \ln(f(x))$ is a concave function.

a) Let $p > 0$ be a fixed, positive real number. Prove that $f(x) = \frac{1}{x^p}$ is logarithmically convex for $x \in (0, \infty)$.

b) Determine whether $f(x) = x^2+1$ is logarithmically convex, logarithmically concave, or neither for $x \in (1, \infty)$.

c) Suppose that we are developing a constant regression model for some data $y_1, y_2, \ldots, y_n \in (1, \infty)$ and we wish to use a loss function of the form:

$$L_{\log}(h, y_i) = 3|y_i - h| - \ln\left(h^3(1 + h^2)\right).$$

Assume that $h > 1$ as well. Prove that the gradient descent algorithm, when applied to the empirical risk function $R_{\log}((y_i)_i, h)$, will converge to a global minimum h^*.

Question 7 Card Combinatorics

A standard deck of cards contains 52 cards. There are 13 cards each in 4 suits (hearts, diamonds, clubs, and spades). Zoe is drawing cards from the deck without replacement.

a) How many different 5-card hands can Zoe draw, regardless of order?

b) How many different 5-card hands can Zoe draw if she must have exactly 2 hearts and 3 diamonds?

c) How many different 5-card hands can Zoe draw if she must have at least one card from each suit? (i.e. 1 heart, 1 diamond, 1 club, and 1 spades, and 1 of any kind)

d) Zoe is playing a card game that requires her to select 4 cards from a standard deck of 52 cards. She must choose:

- 2 red cards (hearts or diamonds)
- 1 black card (spades or clubs)
- 1 face card (Jack, Queen, or King)

However, the red cards **must** be from the hearts suit. Use inclusion-exclusion to calculate the number of valid hands if:

- One of both red cards may be a face card.
- The black card may also be a face card.