PRACTICE FINAL EXAM - DSC 40A, Fall 2025

Full Name:	Solutions
PID:	A12345678
Seat Number:	A1

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.

Solution: Remember that when there is an intercept term (w_0) there needs to be a column of ones.

$$X = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

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.

PID: ____________________

Note:

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

Solution: Recall the equation: $\vec{w} = (X^TX)^{-1}X^T\vec{y}$. To find $\vec{w}$ we need to first multiply the matrix given to us with X^T.

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 1 \end{bmatrix} \tag{1}$$

$$(X^TX)^{-1} \cdot X^T = \begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 6 & 3 & -4 \\ -1 & 3 & 2 & -2 \\ -1 & -4 & -2 & 4 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 1 \end{bmatrix} \tag{2}$$

Here is the first row of our new matrix, follow the same pattern for the rest:

- $3(1) + (-1)(1) + (-1)(1) + (-1)(2) = -1$
- $3(1) + (-1)(0) + (-1)(2) + (-1)(1) = 0$
- $3(1) + (-1)(1) + (-1)(0) + (-1)(1) = 1$
- $3(1) + (-1)(0) + (-1)(1) + (-1)(1) = 1$

Notice how this is just the dot product!

$$(X^TX)^{-1} \cdot X^T = \begin{bmatrix} -1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -2 \\ 0 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Now to find $\vec{w}$ all we need to do is multiply by $\vec{y}$

$$\vec{y} = \begin{bmatrix} -2 \\ 4 \\ 4 \\ 3 \end{bmatrix} \tag{3}$$

$$(X^T X)^{-1} X^T \cdot \vec{y} = \begin{bmatrix} -1 & 0 & 1 & 1 \\ 0 & 1 & 1 & -2 \\ 0 & 1 & 0 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 4 \\ 4 \\ 3 \end{bmatrix} \tag{4}$$

$$\vec{w} = \begin{bmatrix} 9 \\ 2 \\ 1 \\ -7 \end{bmatrix} \tag{5}$$

Notice none of this math was hard, but it is tedious! On the exam we will make sure to give you numbers that can easily be calculated by hand. On the exam, if you write out steps the staff will know what they can give partial credit too!

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!

Solution: Recall $r = y - X\vec{w}$ and MSE $= \frac{1}{n}\sum_{i=1}^{n} r_i^2$.

We can start by calculating $X\vec{w}$.

$$\begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 9 \\ 2 \\ 1 \\ -7 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ 4 \\ 3 \end{bmatrix}$$

Now we can calculate r.

$$r = \begin{bmatrix} -2 \\ 4 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} -2 \\ 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

From here we simply plug this into the Mean Squared Error equation!

$$\text{MSE} = \frac{1}{4}((0)^2 + (0)^2 + (0)^2 + (0)^2) = 0$$

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?)

Solution:

- Let $P(A)$ be the probability of the baked good being a cookie
- Let $P(B)$ be the probability of the baked good being a cake
- Let $P(C)$ be the probability of the baked good being a bread

We can then calculate: $P(A) = \frac{2}{7}$, $P(B) = \frac{2}{7}$, and $P(C) = \frac{3}{7}$.

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.

Solution:

- Let $P(A)$ be the probability of the baked good being a cookie
- Let $P(B)$ be the probability of the baked good being a cake
- Let $P(C)$ be the probability of the baked good being a bread
- Let $P(M)$ be the probability the baked good is Medium Sweetness
- let $P(S)$ be the probability the baked good is soft

To solve this problem we will apply Bayes' Theorem! The equation would be

$$P(\text{Baked Good}|M, S) \propto P(M|\text{Baked Good}) \cdot P(S|\text{Baked Good}) \cdot P(\text{Baked Good})$$

For each baked good we need to find the probability of it having medium sweetness and being soft.

Let's start with **Cookie**:

- There are two instances of cookie
- There are no times the cookie has medium sweetness
- There are no times the cookie is soft

This means: $P(M|A) = 0$ and $P(S|A) = 0$.

Next, let's look at **Cake**:

- There are two instances of cake
- There is one time the cake has medium sweetness
- There are two times the cake is soft

This means: $P(M|B) = \frac{2}{2}$ and $P(S|B) = \frac{2}{2}$.

Finally, let's look at **Bread**:

- There are three instances of bread
- There is one time the bread has medium sweetness
- There are two times the bread is soft

This means $P(M|C) = \frac{1}{3}$ and $P(S|C) = \frac{2}{3}$.

Following this equation:

$$P(\text{Baked Good}|M, S) \propto P(M|\text{Baked Good}) \cdot P(S|\text{Baked Good}) \cdot P(\text{Baked Good})$$

we can find which baked good Sawyer will order!

Cookie: $P(A|M, S) \propto 0 \cdot 0 \cdot \frac{2}{7} = 0$
Cake: $P(B|M, S) \propto \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{7} = \frac{1}{7}$
Bread: $P(C|M, S) \propto \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{3}{7} = \frac{2}{21}$

We can determine Sawyer will most likely order a **cake**!

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

Solution:

- Let $P(A)$ be the probability of the baked good being a cookie
- Let $P(B)$ be the probability of the baked good being a cake
- Let $P(C)$ be the probability of the baked good being a bread
- Let $P(L)$ be the probability the baked good is Medium Sweetness
- let $P(F)$ be the probability the baked good is for a celebration (a fiesta!)

To solve this problem we will apply Bayes' Theorem! The equation would be

$$P(\text{Baked Good}|L, F) \propto P(L|\text{Baked Good}) \cdot P(F|\text{Baked Good}) \cdot P(\text{Baked Good})$$

For each baked good we need to find the probability of it having medium sweetness and being soft. Do not forget to smooth!

Let's start with **Cookie**:

- There are two instances of cookie
- There are no times has low sweetness
- There is one time a cookie is for a celebration

Now we need to do smoothing!

- To do smoothing we will add 1 to the numerator
- To the denominator we need to add the number of possible outcomes!
  - This means because there are 3 kinds of Sweetness levels we add 3 to the denominator for the probability tied to Sweetness.
  - This means because there are 2 kinds of Occasions we add 2 to the denominator for the probability tied to Occasions.

This means our probabilites are $P(L|A) = \frac{0+1}{2+3} = \frac{1}{5}$ and $P(F|A) = \frac{1+1}{2+2} = \frac{1}{2}$

Next, let's look at **Cake**:

- There are two instances of cake
- There are no times the cake has low sweetness
- There is two times the cake is for a celebration

This means: $P(L|B) = \frac{0+1}{2+3} = \frac{1}{5}$ and $P(F|B) = \frac{2+1}{2+2} = \frac{3}{4}$.

Finally, let's look at **Bread**:

- There are three instances of bread
- There is two times the bread has low sweetness
- There is no time the bread is for a celebration

This means $P(L|C) = \frac{2+1}{3+3} = \frac{1}{2}$ and $P(F|C) = \frac{0+1}{3+2} = \frac{1}{5}$.

Following this equation:

$$P(\text{Baked Good}|L, F) \propto P(F|\text{Baked Good}) \cdot P(F|\text{Baked Good}) \cdot P(\text{Baked Good})$$

we can find which baked good Masfiqur will order!

We also need to smooth the probability of selecting each good:

- $P(A) = \frac{2+1}{7+3} = \frac{3}{10}$
- $P(B) = \frac{2+1}{7+3} = \frac{3}{10}$
- $P(C) = \frac{3+1}{7+3} = \frac{4}{10}$

Cookie: $P(A|L, F) \propto \frac{1}{5} \cdot \frac{1}{2} \cdot \frac{3}{10} = \frac{3}{100} = 0.03$
Cake: $P(B|L, F) \propto \frac{1}{5} \cdot \frac{3}{4} \cdot \frac{3}{10} = \frac{9}{200} = 0.03$
Bake: $P(C|L, F) \propto \frac{1}{2} \cdot \frac{1}{5} \cdot \frac{4}{10} = \frac{4}{100} = 0.04$

We can determine Masfiqur is likely to choose a **cake**!

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.

Solution: We use the formula for combinations:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

The total number of ways to select 4 ingredients from 10 is: $\binom{10}{4}$. To find the number of ways that select 4 ingredients that do not include any of the core ingredients we do $10 - 3 = 7$. This means the number of combinations exclusing core ingredients is $\binom{7}{4}$.

What is the probability of having at least 1 core ingredient? $\binom{10}{4} - \binom{7}{4}$

This means the probability of selecting at least one core ingredient is:

$$\frac{\binom{10}{4} - \binom{7}{4}}{\binom{10}{4}}$$

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.

Solution: We use the formula for combinations:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

How many ingredients can Zoe choose if one of them is chocolate chips? $10 - 1 = 9$.

How many ways can Zoe choose 3 ingredients from the remaining 9? $\binom{9}{3}$.

How many ways can Zoe choose 3 ingredients that are not part of the core ingredients? $9 - 3 = 6$, so $\binom{6}{3}$.

The probability that none of the remaining 3 ingredients are core ingredients is $\frac{\binom{6}{3}}{\binom{9}{3}}$.

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.

Solution: We use the formula for combinations:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

The total number of ways Zoe can select 4 ingredients from 10 is: $\binom{10}{4}$.

We can calculate the compliment to find how to choose 4 ingredients from ingredients that are **not** sweet. $10 - 4 = 6$. There are 6 not sweet ingredients so, $\binom{6}{4}$.

This makes the probability of having at least one sweet ingredient: $\frac{\binom{10}{4}-\binom{6}{4}}{\binom{10}{4}}$

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.

Solution: We use the formula for combinations:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

The total number of ways Zoe can select 4 ingredients from 10 is: $\binom{10}{4}$.

How many ways can we choose 2 sweet ingredients out of 4? $\binom{4}{2}$.

The number of ways to choose 2 grains (both flour and oatmeal must be selected) is $\binom{2}{2} = 1$.

The total number of outcomes that have sweet ingredients and 2 grains is: $\binom{4}{2} \cdot \binom{2}{2}$.

This means the probability is: $\frac{\binom{4}{2}\cdot\binom{2}{2}}{\binom{10}{4}}$.

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.

> **Solution:** We have to account for duplicates! In this case we know there are multiples, so we need to account for them!
>
> $$P = \frac{8!}{2! \cdot 2! \cdot 3! \cdot 1!} = 1680$$
>
> Notice: The arrangement for duplicated items will not change the arrangement because the items are identical. This way we do not double count!

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.

> **Solution:** We can use the directions here. We pretend the cookies are a single unit 1!. We then can also subtract out from $n = 8$ the duplicates $(3 - 1 = 2)$. From here we get:
>
> $$P = \frac{6!}{2! \cdot 2! \cdot 1! \cdot 1!} = 180$$

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).

> **Solution:** We need the cakes and cookies to appear before the bread and pies. To solve this, we will first calculate the total number of permutations of all desserts and then determine the number of favorable permutations where cakes and cookies appear before bread and pies.
>
> Total number of permutations There are 8 desserts in total: 2 pies, 2 cakes, 3 cookies, and 1 bread. Since some desserts are repeated, the total number of permutations is the same as what we found in part a:
>
> $$\text{Total permutations} = \frac{8!}{2! \cdot 2! \cdot 3! \cdot 1!} = 1680.$$
>
> To ensure all cakes and cookies appear before bread and pies, we treat the desserts as two groups:
>
> - Group 1: Cakes and cookies (5 items total),
> - Group 2: Bread and pies (3 items total).

Permutations within Group 1 (cakes and cookies):

$$\frac{5!}{2! \cdot 3!} = 10.$$

Permutations within Group 2 (bread and pies):

$$\frac{3!}{2! \cdot 1!} = 3.$$

Since cakes and cookies must appear entirely before bread and pies, there is only one way to order the two groups (Group 1 entirely before Group 2). Thus, the total number of favorable permutations is:

$$\frac{5!}{2! \cdot 3!} \cdot \frac{3!}{2! \cdot 1!} = 10 \cdot 3 = 30.$$

The probability of arranging the desserts such that all cakes and cookies appear before bread and pies is:

$$\text{Probability} = \frac{\text{Favorable permutations}}{\text{Total permutations}} = \frac{\frac{5!}{2! \cdot 3!} \cdot \frac{3!}{2! \cdot 1!}}{\frac{8!}{2! \cdot 2! \cdot 3! \cdot 1!}} = \frac{1}{56}.$$

The probability that cakes and cookies appear before bread and pies is $\frac{1}{56}$.

Question 5 Multiple linear regression

DISCLAIMER This problem was originally going to be on the Fall 2024 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 $\beta_i^{1/2}$.

Solution: We can calculate, using properties from class,

$$\frac{1}{n}\sum_{i=1}^{n} \beta_i |y_i - H(\vec{x}_i)|^2 = \frac{1}{n}\| \begin{bmatrix} \beta_1^{1/2}(y_1 - H(\vec{x}_1)) \\ \beta_2^{1/2}(y_2 - H(\vec{x}_2)) \\ \vdots \\ \beta_n^{1/2}(y_n - H(\vec{x}_n)) \end{bmatrix} \|^2 = \frac{1}{n}\|B^{1/2}(\vec{y} - X^T w)\|^2$$

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* $(\vec{a} - A\vec{z})^{(i)}$.

Solution: According to the hint, we calculate

$$\frac{\partial}{\partial \vec{z}^{(j)}}(\vec{a} - A\vec{z})^{(i)} = \frac{\partial}{\partial \vec{z}^{(j)}}\left(\vec{a}^{(i)} - \sum_{j=1}^{d+1} A_{ij}\vec{z}^{(j)}\right) \tag{6}$$

$$= -A_{ij} \tag{7}$$

since $\vec{a}^{(i)}$ has derivative zero and the only term that depends on $\vec{z}^{(j)}$ is $-A_{ij}\vec{z}^{(j)}$. Therefore

$$\frac{\partial}{\partial \vec{z}^{(j)}}\|\vec{a} - Az\|_2^2 = \frac{\partial}{\partial \vec{z}^{(j)}}\sum_{i=1}^{n}(\vec{a}^{(i)} - (Az)^{(i)})^2 \tag{8}$$

$$= -\sum_{i=1}^{n} 2A_{ij}(\vec{a}^{(i)} - (Az)^{(i)}) \tag{9}$$

$$= -\sum_{i=1}^{n} 2A_{ji}^{T}(\vec{a}^{(i)} - (Az)^{(i)}) \tag{10}$$

$$= -2(A^T(\vec{a} - A\vec{z}))^{(j)} \tag{11}$$

The rest follows.

c) Use the previous part to prove that

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

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

Solution: From part (a), we have

$$R_\beta(w) = \frac{1}{n}\|B^{1/2}(\vec{y} - X\vec{w})\|^2 \tag{12}$$

Using part(b) with $A = B^{1/2}X$ and $\vec{a} = B^{1/2}\vec{y}$, it follows that

$$\nabla R_\beta(w) = -\frac{2}{n}(B^{1/2}X)^T(B^{1/2}\vec{y} - B^{1/2}X\vec{w}) \tag{13}$$

$$= -\frac{2}{n}X^T B^{1/2}(B^{1/2}\vec{y} - B^{1/2}X\vec{w}) \tag{14}$$

$$= -\frac{2}{n}\left(X^T B\vec{y} - X^T BX\vec{w}\right). \tag{15}$$

Therfore since the optimal parameter satisfies $\nabla R_\beta(w^*) = 0$, we have

$$\vec{w}^* = (X^T BX)^{-1}X^T B\vec{y} \tag{16}$$

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.

> **Solution:** There are a few different ways to approach this. First, smaller batches might contribute more uncertainty to the model since they have higher variance in the number of broken toys (a batch of 10 models needs only 1 or 2 broken to have a 10%-20% failure rate), compared to larger batches where the number of broken toys is more reliable. So it might make sense to use weights $\beta_i = 1/n_i$ where n_i is the batch size.
>
> Another reason is that the reliability of the data may vary between batches due to differences in how data is collected. For instance, some batches might have undergone rigorous quality checks (e.g., inspected by multiple workers or using precise automated tools), leading to more accurate measurements of y_i (proportion of broken toys). Other batches might have been assessed using less reliable methods, resulting in noisier estimates of y_i. By assigning higher weights β_i to batches with more reliable measurements (e.g., rigorously inspected batches) and lower weights to noisier batches, the model emphasizes more trustworthy data.
>
> There are lots of additional possibilities.

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 $x \in (0, \infty)$.

Solution: Note that

$$\ln(1/x^p) = \ln(x^{-p}) = -p \ln x$$

And therefore, we have

$$\frac{d^2 f}{dx^2} = \frac{d}{dx} - \frac{p}{x} = \frac{p}{x^2} > 0$$

so by the second derivative test, $\log(1/x^p)$ is convex.

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

Solution:

$$\begin{aligned}
\frac{d}{dx}\log(x^2+1) &= \frac{2x}{x^2+1} \\
\frac{d^2}{dx^2}\log(x^2+1) &= \frac{d}{dx}\frac{2x}{x^2+1} \\
&= \frac{2x^2+2-4x^2}{(x^2+1)^2} \\
&= \frac{2(1-x^2)}{(x^2+1)^2}
\end{aligned}$$

This function is negative whenever $|x| > 1$ and thus $f(x)$ is logarithmically concave on this region.

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^*.

Solution: Notice that

$$\begin{aligned} 3|y_i - h| - \ln\left(h^3(1+h^2)\right) &= 3|y_i - h| - \ln(h^3) - \ln(1+h^2) \\ &= 3|y_i - h| + \ln(1/h^3) - \ln(1+h^2) \end{aligned}$$

We proved in (a) that $\ln(1/h^3)$ is convex, and in (b) that $\ln(1+h^2)$ is concave, and therefore that $-\ln(1+h^2)$ is convex. We know from class that $|y_i - h|$ is convex and thus so is $3|y_i - h|$. Since L_{log} is the sum of convex functions, it must also be convex, so gradient descent is guaranteed to converge.

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?

Solution: To find the number of different 5-card hands Zoe can draw, we use combinations, since the order of the cards does not matter.

The number of ways to choose 5 cards from a deck of 52 cards is given by the combination formula:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Here, $n = 52$ (the total number of cards) and $r = 5$ (the number of cards to choose). Substituting these values:

$$\binom{52}{5} = 2,598,960$$

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

Solution: To solve this, we need to calculate the number of ways to draw 2 hearts from 13 hearts and 3 diamonds from 13 diamonds.

First, we calculate the number of ways to choose 2 hearts from the 13 hearts:

$$\binom{13}{2} = \frac{13 \times 12}{2 \times 1} = 78$$

Next, we calculate the number of ways to choose 3 diamonds from the 13 diamonds:

$$\binom{13}{3} = \frac{13 \times 12 \times 11}{3 \times 2 \times 1} = 286$$

Since the selection of hearts and diamonds are independent, we multiply the two combinations together to find the total number of ways to draw the 5 cards:

$$\text{Total ways} = \binom{13}{2} \times \binom{13}{3} = 78 \times 286 = 22,308$$

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)

Solution: To calculate the number of 5-card hands Zoe can draw where there is at least one card from each suit (hearts, diamonds, clubs, and spades), we must ensure that the hand satisfies the following criteria:

- The hand contains exactly one card from three of the suits.
- The hand contains two cards from the remaining suit.

Since there are four suits (hearts, diamonds, clubs, spades), we first choose the suit that contributes two cards. There are: 4 ways to do this because there are 4 suits.

Once the suit with two cards is chosen, we need to select 2 cards from the 13 cards in that suit. The number of ways to do this is: $\binom{13}{2}$

For the remaining three suits, we must select exactly one card from each suit. Since there are 13 cards in each suit, the number of ways to select one card from a single suit is 13. Thus, the total number of ways to select one card from each of the three suits is: 13^3.

The total number of 5-card hands that meet the given criteria is the product of all these choices:

$$4 \cdot \binom{13}{2} \cdot 13^3.$$

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.

Solution: We calculate the number of valid hands by using inclusion-exclusion.

First, we compute the total ways to choose cards ignoring any overlap between face cards and the suit or color requirements.

Choose 2 red cards from hearts (13 cards in the hearts suit):

$$\binom{13}{2}.$$

Choose 1 black card (26 black cards total):

$$\binom{26}{1}.$$

Choose 1 face card (12 face cards total across all suits):

$$\binom{12}{1}.$$

Thus, the initial total is:

$$\binom{13}{2} \cdot \binom{26}{1} \cdot \binom{12}{1}.$$

We subtract cases where face cards violate suit or color restrictions:

Case A: A red face card is chosen as one of the red cards.
Red face cards come only from hearts (3 cards: Jack, Queen, King). The number of ways to include one red face card in the 2 red cards is:

$$\binom{3}{1} \cdot \binom{10}{1},$$

where $\binom{3}{1}$ selects 1 red face card, and $\binom{10}{1}$ selects the remaining red card from the non-face red cards in hearts.

Case B: The black card is also a face card. There are 6 black face cards (3 each in spades and clubs). The number of ways to choose a black face card is:

$$\binom{6}{1}.$$

We add back cases where a single card violates both conditions (a red face card is selected, and the black card is also a face card), because they were subtracted twice:
Case C: A red face card and a black face card are both selected. There are 3 red face cards and 6 black face cards, so the number of ways to choose such a pair is:

$$\binom{3}{1} \cdot \binom{6}{1}.$$

Using inclusion-exclusion, the **total number of valid hands** is:

$$\binom{13}{2}\cdot\binom{26}{1}\cdot\binom{12}{1}-\left[\binom{3}{1} \cdot \binom{10}{1} \cdot \binom{26}{1}\right]-\left[\binom{6}{1} \cdot \binom{13}{2} \cdot \binom{12}{1}\right]+\left[\binom{3}{1} \cdot \binom{10}{1} \cdot \binom{6}{1}\right].$$