Final Part I - DSC 40A, Winter 2024

Instructions

- This is a 50-minute exam consisting of 4 questions worth a total of 40 points.
- The only allowed resource is the unlimited pages of hand-written notes.
- No calculators.
- Please write neatly and stay within the provided boxes.
- You may fill out the **front page only** until you are instructed to start.

Statement of Academic Integrity

By submitting your exam, you are attesting to the following statement of academic integrity.

I will act with honesty and integrity during this exam.

Name:	Solutions
PID:	A12345678
Seat you are in:	

Version - A

1. (8 points) Suppose there is a dataset containing 10,000 integers, 2,500 of them are 3, 2,500 of them are 5, 4,500 of them are 7, and 500 of them are 9.

a) (4 points) Calculate the median of this dataset.

6

b) (4 points) How does the mean of this dataset compared to its median?

○ The mean is larger than the median

● The mean is smaller than the median

○ The mean and the median are equal

2. (8 points) You and a friend independently perform gradient descent on the same function, but after 200 iterations, you have different results. Which of the following is sufficient **on its own** to explain the difference in your results? **Select all that apply.**

☐ The function is nonconvex.

☐ The function is not differentiable.

■ You and your friend chose different learning rates.

☐ You and your friend chose the same initial predictions.

☐ None of the above.

3. (10 points) 3. Suppose there is a dataset contains 4 values: -2, -1, 2, 4. You would like to use gradient descent to minimize the mean square error over this dataset.

a) (2 points) Write down the expression of mean square error and its derivative given this dataset

$R_{sq}(h) =$ $\frac{1}{4}\sum_{i=1}^{4}(y_i - h)^2$

$\frac{dR_{sq}(h)}{dh} =$ $\frac{1}{2}\sum_{i=1}^{4}(h - y_i)$

b) (4 points) Suppose you choose the initial position to be h_0 and the learning rate to be $\frac{1}{4}$. After two gradient descent steps, $h_2 = \frac{1}{4}$. What is the value of h_0?

$h_0 =$

Solution: The gradient descent equation is given by:

$$h_i = h_{i-1} - \alpha\frac{dR_{sq}(h_{i-1})}{dh}$$

We can plug in $\alpha = 1/4$, $h_2 = 1/4$, and $i = 2$, in that case we obtain:

$$\begin{aligned}
\frac{1}{4} &= h_1 - \alpha \frac{dR_{sq}(h_1)}{dh} \\
&= h_1 - \alpha \frac{1}{2} \sum_{i=1}^{4} (h_1 - y_i) \\
&= h_1 - \frac{1}{8}(h_1 + 2 + h_1 + 1 + h_1 - 2 + h_1 - 4) \\
&= h_1 - \frac{1}{8}(4h_1 - 3)
\end{aligned}$$

Solving this equation, we obtain that $h_1 = -\frac{1}{4}$. We can then repeat this step once more to obtain h_0:

$$\begin{aligned}
-\frac{1}{4} &= h_0 - \alpha \frac{dR_{sq}(h_0)}{dh} \\
&= h_0 - \frac{1}{4}(h_0 + 2 + h_0 + 1 + h_0 - 2 + h_0 - 4) \\
&= h_0 - \frac{1}{8}(4h_0 - 3)
\end{aligned}$$

Solving this equation, we obtain that $h_0 = -\frac{5}{4}$.

c) (4 points) Given that we set the tolerance of gradient descent to be 0.1. How many **additional steps** beyond h_2 do we need to take to reach convergence?

○ 0 ○ 1 ● 2 ● 3 ○ 4 ○ It will never converge

Solution: The gradient descent equation is given by:

$$h_i = h_{i-1} - \alpha \frac{dR_{sq}(h_{i-1})}{dh}$$

Now we start from $\alpha = 1/4$, $h_2 = 1/4$, and $i = 2$, in that case we obtain:

$$\begin{aligned}
h_3 &= h_2 - \alpha \frac{dR_{sq}(h_2)}{dh} \\
&= h_2 - \frac{1}{8}(4h_2 - 3) \\
&= \frac{1}{4} - \frac{1}{8}(1 - 3) \\
&= \frac{1}{2}
\end{aligned}$$

Iteratively, we have:

$$\begin{aligned} h_4 &= h_3 - \alpha \frac{dR_{sq}(h_3)}{dh} \\ &= h_3 - \frac{1}{8}(4h_3 - 3) \\ &= \frac{1}{2} - \frac{1}{8}(2-3) \\ &= \frac{5}{8} \end{aligned}$$

$$\begin{aligned} h_5 &= h_4 - \alpha \frac{dR_{sq}(h_4)}{dh} \\ &= h_3 - \frac{1}{8}(4h_4 - 3) \\ &= \frac{5}{8} - \frac{1}{8}(\frac{5}{2} - 3) \\ &= \frac{11}{16} \end{aligned}$$

from h_4 to h_5, the change is smaller than the tolerance. That means we need additional 2 or 3 steps to reach convergence (depending on if you actually perform h_4 to h_5, so both 2 and 3 are considered correct answer.

4. (14 points) Albert collected 400 data points from a radiation detector. Each data point contains 3 feature: feature A, feature B and feature C, along with the true particle energy E. Albert want to design a linear regression algorithm to predict the energy E of each particle, given one or a combination(s) of feature A, B, C. As the first step, Albert calculated the correlation coefficients among A, B, C and E. He wrote it down in the following table, where each cell of the table represents the correlaton of two terms:

	Feature A	**Feature B**	**Feature C**	**Energy E**
Feature A	1	-0.99	0.13	0.8
Feature B	-0.99	1	0.25	-0.95
Feature C	0.13	0.25	1	0.72
Energy E	0.8	-0.95	0.72	1

a) (4pt) Albert want to start with a simple model: fitting only a single feature to obtain the true energy (i.e. $y = w_0 + w_1 x$). Which feature should he choose as x to get the lowest mean square error?

○ A ● B ○ C

Solution: B is the correct answer, because it has the highest absolute correlation (0.95), the negative sign in front of B just means it is negatively correlated to energy and it can be compensated by a negative sign in the weight.

b) (2pt) Albert want to add another feature into his linear regression in part a) to further boost the model's performance. (i.e. $y = w_0 + w_1 x + +w_2 x_2$) Which feature should he choose as x_2 to make additional improvements?

○ A ○ B ● C

Solution: C is the correct answer, because although A has a higher correlation with energy, it also has an extremely high correlation with B (-0.99), that means adding A into the fit will not be too much useful, since it provides almost the same information as B.

Albert further refine his algorithm by fitting a prediction rule of the form:

$$H(A, B, C) = w_0 + w_1 \cdot A \cdot C + w_2 \cdot B^{C-7}$$

c) (3pt)Given this prediction rule, What are the dimensions of the design matrix X?

$\boxed{400}$ rows $\times$ $\boxed{3}$ columns

d) (5pt) Now Albert solve the normal equations and find the solution to be:

$$\vec{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \\ w_2^* \end{bmatrix}$$

To improve on this result, Albert decides to modify his design matrix with the following steps:

1. Add 1 to every entry to the first column
2. Interchange the second and the third column

Let X_a be the modified design matrix. Let $\vec{w_a}^* = (X_a^T X_a)^{-1} X_a^T \vec{y}$. Express the components $\vec{w_a}^*$ in terms of w_0^*, w_1^*, w_2^*, which were the components of $\vec{w}^*$.

$$\vec{w_a}^* = \begin{bmatrix} \\ \\ \\ \end{bmatrix}$$

Solution:

$$\vec{w}^* = \begin{bmatrix} w_0^*/2 \\ w_2^* \\ w_1^* \end{bmatrix}$$