Midterm Exam - DSC 40A, Spring 2024

Full Name:

PID:

Seat Number:

Instructions:

- This exam consists of 5 questions, worth a total of 64 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.
  - When asked to do so, please place your final answer in a $\boxed{\text{box}}$.
  - ◯ In multiple choice questions, select only one option and completely fill in the corresponding bubble — if we cannot tell which option you selected, you may not receive credit.
- You may refer to a single two-sided index card of maximum size 4 inches by 6 inches that you wrote on by hand by yourself. 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 (14 pts)

Vectors get lonely, and so we will give each vector one friend to keep them company. Specifically, if $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$, $\vec{v_f}$ is the friend of $\vec{v}$, where $\vec{v_f} = \begin{bmatrix} -v_2 \\ v_1 \end{bmatrix}$.

a) (2 pts) Prove that $\vec{v}$ and $\vec{v_f}$ are orthogonal.

Consider the vectors $\vec{c}$ and $\vec{d}$, defined below.

$$\vec{c} = \begin{bmatrix} 1 \\ 7 \end{bmatrix} \qquad \vec{d} = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

The next few parts ask you to write various vectors as scalar multiples of either $\vec{c}$, $\vec{c_f}$, $\vec{d}$, or $\vec{d_f}$, where $\vec{c_f}$ and $\vec{d_f}$ are the friends of $\vec{c}$ and $\vec{d}$, respectively. In each part, write **one number** in the box, and fill in **one bubble**. Part (b) has already been done for you.

b) A vector in span($\vec{d}$) that is twice as long as $\vec{d}$.

2 $\times$ ○ $\vec{c}$ ○ $\vec{c_f}$ ● $\vec{d}$ ○ $\vec{d_f}$

c) (3 pts) The projection of $\vec{c}$ onto span($\vec{d}$).

$\times$ ○ $\vec{c}$ ○ $\vec{c_f}$ ○ $\vec{d}$ ○ $\vec{d_f}$

d) (3 pts) The error vector of the projection of $\vec{c}$ onto span($\vec{d}$).

$\times$ ○ $\vec{c}$ ○ $\vec{c_f}$ ○ $\vec{d}$ ○ $\vec{d_f}$

e) (3 pts) The projection of $\vec{d}$ onto span($\vec{c}$).

$\times$ ○ $\vec{c}$ ○ $\vec{c_f}$ ○ $\vec{d}$ ○ $\vec{d_f}$

f) (3 pts) The error vector of the projection of $\vec{d}$ onto span($\vec{c}$).

$\times$ ○ $\vec{c}$ ○ $\vec{c_f}$ ○ $\vec{d}$ ○ $\vec{d_f}$

Question 2 (11 pts)

Consider a dataset of n values, $y_1, y_2, ..., y_n$, all of which are non-negative. We're interested in fitting a constant model, $H(x) = h$, to the data, using the new "Sun God" loss function:

$$L_{\text{sungod}}(y_i, h) = w_i \left(y_i^2 - h^2\right)^2$$

Here, w_i corresponds to the "weight" assigned to the data point y_i, the idea being that different data points can be weighted differently when finding the optimal constant prediction, h^*.

For example, for the dataset $y_1 = 1, y_2 = 5, y_3 = 2$, we will end up with different values of h^* when we use the weights $w_1 = w_2 = w_3 = 1$ and when we use weights $w_1 = 8, w_2 = 4, w_3 = 3$.

a) (3 pts) Find $\frac{\partial L_{\text{sungod}}}{\partial h}$, the derivative of the Sun God loss function with respect to h. Show your work, and put a $\boxed{\text{box}}$ around your final answer.

b) (6 pts) Prove that the constant prediction that minimizes empirical risk for the Sun God loss function is:

$$h^* = \sqrt{\frac{\sum_{i=1}^n w_i y_i^2}{\sum_{i=1}^n w_i}}$$

c) (2 pts) For a dataset of non-negative values $y_1, y_2, ..., y_n$ with weights $w_1, 1, ..., 1$, evaluate:

$$\lim_{w_1 \to \infty} h^*$$

- ◯ The maximum of $y_1, y_2, ..., y_n$
- ◯ The mean of $y_1, y_2, ..., y_{n-1}$
- ◯ The mean of $y_2, y_3, ..., y_n$
- ◯ The mean of $y_2, y_3, ..., y_n$, multiplied by $\frac{n}{n-1}$
- ◯ y_1
- ◯ y_n

PID: ______________________________

Question 3 (12 pts)

Consider a dataset of n values, $y_1, y_2, ..., y_n$, where $y_1 < y_2 < ... < y_n$. Let $R_{\text{abs}}(h)$ be the mean absolute error of a constant prediction h on this dataset of n values.

Suppose that we introduce a new value to the dataset, α. Let $S_{\text{abs}}(h)$ be the mean absolute error of a constant prediction h on this new dataset of $n + 1$ values.

We're given that:

- $n > 5$.
- α is not equal to any of $y_1, y_2, ..., y_n$.
- All values of h between 7 and 9 minimize $S_{\text{abs}}(h)$.
- The slope of $S_{\text{abs}}(h)$ on the line segment immediately to the right of α is $\frac{5-n}{1+n}$.

a) (2 pts) In the problem statement, we were told that "all values between 7 and 9 minimize $S_{\text{abs}}(h)$." More specifically, what interval of values h minimize $S_{\text{abs}}(h)$?

◯ $7 < h < 9$ ◯ $7 \leq h < 9$ ◯ $7 < h \leq 9$ ◯ $7 \leq h \leq 9$

b) (4 pts) Which value(s) minimize $R_{\text{abs}}(h)$? Give your answer(s) as integer(s) with no variables. Show your work, and put a $\boxed{\text{box}}$ around your final answer(s).

Hint: Don't start by trying to expand $\frac{1}{n}\sum_{i=1}^{n} |y_i - h|$ — instead, think about what removing α does.

c) (6 pts) What is the slope of $S_{\text{abs}}(h)$ on the line segment immediately to the left of α? Give your answer in the form of an expression involving n. Show your work, and put a $\boxed{\text{box}}$ around your final answer.

PID: ______________________________

Question 4 (19 pts)

Suppose we want to fit a hypothesis function of the form:

$$H(x) = w_0 + w_1 x^2$$

Note that this is *not* the simple linear regression hypothesis function, $H(x) = w_0 + w_1 x$.

To do so, we will find the optimal parameter vector $\vec{w}^* = \begin{bmatrix} w_0^* \\ w_1^* \end{bmatrix}$ that satisfies the normal equations. The first 5 rows of our dataset are as follows, though note that our dataset has n rows in total.

x	y
2	4
-1	4
3	4
-7	4
3	4

Suppose that $x_1, x_2, ..., x_n$ have a mean of $\bar{x} = 2$ and a variance of $\sigma_x^2 = 10$.

a) (3 pts) Write out the first 5 rows of the design matrix, X.

b) (3 pts) Suppose, just in part (b), that after solving the normal equations, we find $\vec{w}^* = \begin{bmatrix} 2 \\ -5 \end{bmatrix}$. What is the predicted y value for the augmented feature vector $\text{Aug}(\vec{x}) = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$? Give your answer as an integer with no variables. Show your work, and put a $\boxed{\text{box}}$ around your final answer.

c) (5 pts) Let $X_{\text{tri}} = 3X$. Using the fact that $\sum_{i=1}^{n} x_i^2 = n\sigma_x^2 + n\bar{x}^2$, determine the value of the bottom-left value in the matrix $X_{\text{tri}}^T X_{\text{tri}}$, i.e. the value in the second row and first column. Give your answer as an expression involving n. Show your work, and put a $\boxed{\text{box}}$ around your final answer.

d) (8 pts) Consider the following four hypothesis functions:

- $H_1(x) = H(x) = w_0 + w_1 x^2$
- $H_2(x) = w_0$
- $H_3(x) = w_0 + w_1 x$
- $H_4(x) = w_0 + w_1 x + w_2 x^2$

Let H_1^*, H_2^*, H_3^*, and H_4^* be the versions of all four hypothesis functions that are using optimal parameters. In the subparts below, fill in the blanks.

(i) The mean squared error of H_1^* is ____ the mean squared error of H_2^*.

◯ greater than ◯ greater than or equal to ◯ equal to
◯ less than ◯ less than or equal to ◯ impossible to tell

(ii) The mean squared error of H_1^* is ____ the mean squared error of H_3^*.

◯ greater than ◯ greater than or equal to ◯ equal to
◯ less than ◯ less than or equal to ◯ impossible to tell

(iii) The mean squared error of H_1^* is ____ the mean squared error of H_4^*.

◯ greater than ◯ greater than or equal to ◯ equal to
◯ less than ◯ less than or equal to ◯ impossible to tell

(iv) In ____ of the hypothesis functions H_1^*,H_2^*, H_3^*, and H_4^*, the sum of the residuals of the function's predictions is 0.

◯ none ◯ 1 ◯ 2 ◯ 3 ◯ all 4

Question 5 (8 pts)

Consider a dataset of n points, $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ where:

- $x_1 < x_2 < ... < x_n$, and $x_1, x_2, ..., x_n$ have a variance of $\sigma_x^2 = 15$ and a range of 20 (the range of a collection of values is the difference between the largest and smallest value).
- $y_1 > y_2 > ... > y_n$, and $y_1, y_2, ..., y_n$ have a variance of $\sigma_y^2 = 8$ and a range of 6.

We fit two linear hypothesis functions using squared loss:

- One hypothesis function is fit with a "swapped" version of the dataset, where x_1 and x_n are swapped — that is, it uses the dataset $(x_n, y_1), (x_2, y_2), ..., (x_{n-1}, y_{n-1}), (x_1, y_n)$. Note that only two of the points in this dataset are different than in the original dataset. We'll call the optimal slope and intercept of this hypothesis function w_1^{swap} and w_0^{swap}, respectively.
- Another hypothesis function is fit with the original dataset, $(x_1, y_1), (x_2, y_2), ..., (x_{n-1}, y_{n-1}), (x_n, y_n)$. We'll call the optimal slope and intercept of this hypothesis function w_1^{orig} and w_0^{orig}, respectively.

On the next page, in the space provided, prove that:

$$|w_1^{\text{swap}} - w_1^{\text{orig}}| = \frac{8}{n}$$

Hint: Approach this problem similarly to Problem 2 on Homework 3 ("Shout for Stroud"). Also, think about how you can express $\sum_{i=1}^{n}(x_i - \bar{x})^2$ in terms of n and σ_x^2.

Feel free to use the space here for scratch work, but we will only grade what appears in the box on the next page.

Put your proof for Question 5 in the box below.

PID: ______________________

Make sure you've written your PID in the space provided in the top right corner of every page of this exam.

Feel free to draw us a picture about DSC 40A below.

And here's a free point!