Data 100, Spring 2025

Homework 1 Math Prerequisites

Due Date: Thursday, January 30, 11:59 PM PT

Submission Instructions

You must submit this assignment to Gradescope by the on-time deadline, **Thursday, January 30, 11:59 PM PT**. Please read the syllabus for **the Slip Day policy**. No late submissions beyond the Slip Day policy will be accepted unless additional accommodations have been arranged prior. While course staff is happy to help you if you encounter difficulties with submission, we may not be able to respond to last-minute requests for assistance (TAs need to sleep, after all!). **We strongly encourage you to plan to submit your work to Gradescope several hours before the stated deadline.** This way, you will have ample time to contact staff for submission support.

There are four parts to this assignment listed on Gradescope:

- **Homework 1 Coding**: Submit your Jupyter Notebook zip file for Homework 1A, which can be generated and downloaded from DataHub using the grader.export() cell provided.

- **Homework 1 Coding Written**: Gradescope will automatically submit the PDF from the zip file submitted earlier. You do not need to submit anything to this assignment yourself, but **you are responsible** for checking that it submitted properly.

- **Homework 1 Math Prerequisites**: Submit a PDF to Gradescope that contains all your answers to all questions in Homework 1 (Math Prerequisites).

- Syllabus Quiz: The assignment is multiple-choice style on Gradescope. You may change or update your answers anytime before the deadline.

To receive credit on this assignment, **you must submit both your coding and written portions to their respective Gradescope portals as well as the syllabus quiz**.

You can answer the below Homework 1 math prerequisite written questions in one of many ways:

1. Type your answers. We recommend LaTeX, the math typesetting language. Overleaf is a great tool to type in LaTeX.

2. Download this PDF, print it out, and write directly on these pages. If you have a tablet, you may save this PDF and write directly on it.

3. Write your answers on a blank sheet of physical or digital paper. Note: If you write your answers on physical paper, use a scanning application (e.g., CamScanner, Apple Notes) to generate a PDF.

Important: When submitting Homework 1 (Math Prerequisites) on Gradescope, you **must tag pages to each question correctly** (it prompts you to do this after submitting your work). This significantly streamlines the grading process for our readers. Failure to do this may result in a score of 0 for untagged questions.

You are responsible for ensuring your submission follows our requirements and that the automatic submission for Homework 1 Coding Written answers went through properly. We will not be granting regrade requests nor extensions to submissions that don't follow instructions. If you encounter any difficulties with submission, please don't hesitate to contact staff before the deadline.

Collaboration Policy

Data science is a collaborative activity. While you may talk with others about the homework, we ask that you write your solutions individually. If you discuss the assignments with others, please include their names below.

Colin Rondon

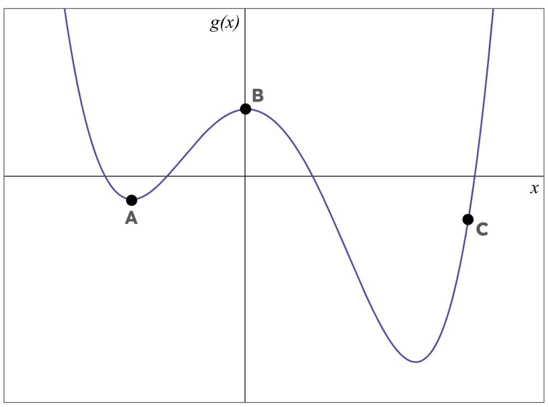
Homework 1 Manually Graded Questions

- This is not a question. This is a reminder to check and make sure that your Homework 1 manually graded questions (automatically submitted as a PDF by Gradescope when the HW 1 Coding assignment is submitted) were properly generated and uploaded.

Note: If you are looking for some resources to supplement your Calculus and Linear Algebra knowledge you can reference the resources linked on the course website.

Calculus and Algebra

1. (3 points) For **each point**, A, B, and C, on the function $g(x)$ plotted below, describe the values of $\frac{dg}{dx}$ and $\frac{d^2g}{dx^2}$ by comparison to 0. For example, $\frac{dg}{dx} < 0$, $\frac{dg}{dx} > 0$, or $\frac{dg}{dx} = 0$.



	A	B	C
$\frac{dg}{dx}$	$=0$	$=0$	>0
$\frac{d^2g}{dx^2}$	>0	<0	>0

2. (5 points) Let $f(x) = \frac{x^3}{3} - x^2 - 3x + 7$. Find the local minimum and maximum point(s) of $f(x)$, and write them in the form (a, b), specifying whether each point is a minimum or maximum. **Coordinates should be kept in fractions.**

   Additionally, provide in your answer if $f(x)$ has an absolute minimum or maximum over its entire domain with their corresponding values. Otherwise, state that there is no absolute maximum or minimum. As a reminder, ∞ and $-\infty$ are not considered absolute maxima and minima respectively.

$f'(x) = x^2 - 2x - 3$

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

$x = 3, -1$

$f''(x) = 2x - 2$

absolute extrema check:

$\lim_{x \to \infty} \frac{x^3}{3} - x^2 - 3x + 7 = \infty$

$\lim_{x \to -\infty} \frac{x^3}{3} - x^2 - 3x + 7 = -\infty$

no absolute extrema.

$\underline{x = 3}$

$f(3) = 9 - 9 - 9 + 7 = -2$

$\underline{x = -1}$

$f(-1) = -\frac{1}{3} - 1 + 3 + 7 = -\frac{1}{3} + \overset{\frac{27}{3}}{9} = \frac{26}{3}$

crit points: $(3, -2)$, $(-1, \frac{26}{3})$

$f''(3) = 6 - 2 = 4$

$f''(-1) = -2 - 2 = -4$

local minimum at $(3, -2)$

local maximum at $(-1, \frac{26}{3})$

$$-\frac{\ln x}{y^7-4z} + \frac{z}{y^7-4z} + \frac{3x^2 z}{y^4} - e^{2xy}\ln z + 10y^2 z$$

3. (3 points) Let $h(x,y,z) = -\frac{ln(x)-z}{y^7-4z} + \frac{3x^2z}{y^4} - e^{2xy}ln(z) + 10y^2z$.

   (a) Holding all other variables constant, take the partial derivative of $h(x,y,z)$ with respect to x, $\frac{\partial}{\partial x}h(x,y,z)$.

$$\boxed{h_x = -\frac{1}{x(y^7-4z)} + \frac{6xz}{y^4} - 2y e^{2xy}\ln z + 10y^2 z}$$

   (b) Holding all other variables constant, take the partial derivative of $h(x,y,z)$ with respect to y, $\frac{\partial}{\partial y}h(x,y,z)$.

$$h_y = -(\ln x - z)\frac{\partial}{\partial y}(y^7-4z)^{-1} + (3x^2z)\frac{\partial}{\partial y}(y^{-4}) - 2e^{2xy}\ln z + 20yz$$

$$= -(\ln x - z)(-(y^7-4z)^{-2})(7y^6) + (3x^2z)(-4y^{-5}) - 2xe^{2xy}\ln z + 20yz$$

$$\boxed{= \frac{(7y^6)(\ln x - z)}{(y^7-4z)^2} - \frac{12x^2z}{y^5} - 2xe^{2xy}\ln z + 20yz}$$

Probability and Statistics

4. (8 points) Much of data analysis involves interpreting proportions – lots and lots of related proportions. So let's recall the basics. It might help to start by reviewing the main rules from Data 8 (Chapter 9.5), with particular attention to what's being multiplied in the multiplication rule.

   For this question, assume a specific section of Doe Library features 32 recently published books, divided into four genres: 9 Contemporary novels, 8 Thrillers, 7 Sci-Fi books, and 8 Fantasy novels. The books are classified into four distinct *genres* (contemporary, thriller, sci-fi, and fantasy). The average number of pages in each genre is provided in the table below.

   Note: You may leave your solutions in equation form, and values can remain as proportions.

Genre	Number of Books	Average Number of Pages
Contemporary	9	320
Thriller	8	280
Sci-Fi	7	280
Fantasy	8	300
Total	**32**	—

   (a) Xiaorui selects 2 books to borrow at random, without replacement. What is the probability that the genre of the first book is Contemporary and the second book is of a different genre?

$$\frac{9}{32} \cdot \frac{23}{31}$$

(b) Sam borrows one book at random on Wednesday and returns it on Friday. On Saturday, Xiaorui also borrows one book at random. What is the probability that **at least one of the books borrowed by Sam and Xiaorui is a Sci-Fi book?**

$$1-\left(\frac{25}{32}\right)^2$$

(c) If a book is to be selected at random, what is the probability that the selected book is either a Thriller or belongs in the category of an average page count of at least 300?

≥ 300

contemp

fantasy

$$\frac{8}{32}+\frac{9}{32}+\frac{8}{32}$$

(d) If a book is to be selected at random, what is the probability that the selected book is either a Fantasy or belongs in the category of an average page count of at least 300?

$$\frac{8}{32}+\frac{9}{32}$$

Linear Algebra

5. (6 points) A common representation of data uses matrices and vectors, so it is helpful to familiarize ourselves with linear algebra notation, as well as some simple operations.

   Define a vector $\vec{v}$ to be a column vector. Then, the following properties hold:

   - $c\vec{v}$ with c some constant, is equal to a new vector where every element in $c\vec{v}$ is equal to the corresponding element in $\vec{v}$ multiplied by c. For example, $2\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}2\\4\end{bmatrix}$.
   - $\vec{v}_1 + \vec{v}_2$ is equal to a new vector with elements equal to the elementwise addition of $\vec{v}_1$ and $\vec{v}_2$. For example, $\begin{bmatrix}1\\2\end{bmatrix} + \begin{bmatrix}-3\\4\end{bmatrix} = \begin{bmatrix}-2\\6\end{bmatrix}$.

   The above properties form our definition for a **linear combination** of vectors. $\vec{v}_3$ is a linear combination of $\vec{v}_1$ and $\vec{v}_2$ if $\vec{v}_3 = a\vec{v}_1 + b\vec{v}_2$, where a and b are some constants.

   Oftentimes, we stack column vectors to form a matrix. Define the **column rank** of a matrix A to be equal to the maximal number of linearly independent columns in A. A set of columns is **linearly independent** if no column can be written as a linear combination of any other column(s) within the set. If all columns in a matrix are linearly independent, it means that the matrix is **full column rank**.

   For example, let A be a matrix with 4 columns. If three of these columns are linearly independent, but the fourth can be written as a linear combination of the other three, then $\text{rank}(A) = 3$. Alternatively, if all four columns of A were linearly independent, $\text{rank}(A) = 4$, and A would be full column rank.

   For each of the following matrices, state the rank of the matrix **and** whether or not the matrix is full column rank. If the matrix is not full column rank, **also** give a linear relationship among the vectors—for example: $\vec{v}_1 = \vec{v}_2$.

   (a) $\vec{v}_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix}1\\1\\1\end{bmatrix}$, $A = \begin{bmatrix}| & |\\ \vec{v}_1 & \vec{v}_2\\ | & |\end{bmatrix}$

   $A = \begin{bmatrix}1 & 1\\0 & 1\\0 & 1\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 1\\0 & 0\end{bmatrix}$

Rank$(A) = 2$, full column rank

(b) $\vec{v}_1 = \begin{bmatrix}-4\\-3\end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix}4\\3\end{bmatrix}$, $B = \begin{bmatrix}| & |\\ \vec{v}_1 & \vec{v}_2\\ | & |\end{bmatrix}$

$$B = \begin{bmatrix}-4 & 4\\ -3 & 3\end{bmatrix} = \begin{bmatrix}1 & -1\\ 0 & 0\end{bmatrix}$$

$R_2 \to -\frac{3}{4}R_1 + R_2$

$\vec{v}_1 = -\vec{v}_2$

Rank $(B) = 1$, no full column rank

(c) $\vec{v}_1 = \begin{bmatrix}0\\1\end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix}5\\0\end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix}15\\15\end{bmatrix}$, $C = \begin{bmatrix}| & | & |\\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3\\ | & | & |\end{bmatrix}$

$$C = \begin{bmatrix}0 & 5 & 15\\ 1 & 0 & 15\end{bmatrix} = \begin{bmatrix}1 & 0 & 15\\ 0 & 5 & 15\end{bmatrix}$$

Rank $(C) = 2$, full column rank

(d) $\vec{v}_1 = \begin{bmatrix}-2\\-2\\5\end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix}2\\4\\-2\end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix}0\\2\\3\end{bmatrix}$, $D = \begin{bmatrix}| & | & |\\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3\\ | & | & |\end{bmatrix}$

$$D = \begin{bmatrix}-2 & 2 & 0\\ -2 & 4 & 2\\ 5 & -2 & 3\end{bmatrix} = \begin{bmatrix}-2 & 2 & 0\\ 0 & 2 & 2\\ 0 & 3 & 3\end{bmatrix} = \begin{bmatrix}1 & -1 & 0\\ 0 & 1 & 1\\ 0 & 0 & 0\end{bmatrix}$$

$R_2 \to -R_1 + R_2$

$R_3 \to \frac{5}{2}R_1 + R_3$

$R_3 \to -\frac{3}{2}R_2 + R_3$

$R_1 \to -\frac{1}{2}R_1$

$R_2 \to \frac{1}{2}R_2$

Rank $(D) = 2$, no full column rank

$v_3 = v_1 + v_2$