

Assignment #9

Due: 6:00pm Wednesday 29 November 2023

Upload at: <https://www.gradescope.com/courses/606160/assignments/3650401>

Assignments in COS 302 should be done individually. See the [course syllabus](#) for the collaboration policy.

Remember to append your Colab PDF as explained in the first homework, with all outputs visible.
When you print to PDF it may be helpful to scale at 95% or so to get everything on the page.

Problem 1 (18pts)

Consider the univariate function

$$f(x) = x^3 + 6x^2 - 3x - 5.$$

Find its stationary points and indicate whether they are maxima, minima, or saddle points (point of inflection).

Problem 2 (20pts)

Consider the following scalar-valued function

$$f(x, y, z) = x^2y + \sin(z + 6y).$$

- (A) Compute partial derivatives with respect to x , y , and z .
- (B) We can consider f to be a function that takes a vector $\boldsymbol{\theta} \in \mathbb{R}^3$ as input, where $\boldsymbol{\theta} = [x, y, z]^T$. Write the gradient as a vector and evaluate it at $\boldsymbol{\theta} = [3, \pi/2, 0]^T$.

Problem 3 (20pts)

The purpose of this problem is to demonstrate Clairaut's Theorem, which states that in general, the order in which you perform partial differentiation does not matter. (Technically there are assumptions that the function must satisfy, but they are almost always true for the kinds of functions we care about in machine learning.) Consider the following scalar-valued function,

$$f(x, y, z) = x \sin(xy),$$

where $x, y \in \mathbb{R}$.

(A) Compute $\frac{\partial}{\partial x} \frac{\partial}{\partial y} f(x, y)$. This means we first compute the partial derivative of f with respect to y , then compute the partial derivative of the resulting function with respect to x . This is sometimes denoted $\partial_{xy} f$.

(B) Compute $\frac{\partial}{\partial y} \frac{\partial}{\partial x} f(x, y)$.

You should have gotten the same answer for parts (A) and (B), which demonstrates Clairaut's Theorem. This holds more generally for n variables as well, that is, if $f(x_1, x_2, \dots, x_n)$ is a function of n variables. This result can be useful when differentiating functions, since it's possible that it's much more convenient computationally-wise to differentiate in an order that you choose.

Problem 4 (20pts)

Consider the following vector function from \mathbb{R}^3 to \mathbb{R}^3 :

$$f(\mathbf{x}) = \begin{bmatrix} \sin(x_1 x_2 x_3) \\ \cos(x_2 + x_3) \\ \exp\{-\frac{1}{2}(x_3^2)\} \end{bmatrix}$$

- (A) What is the Jacobian matrix of $f(\mathbf{x})$?
- (B) Write the determinant of this Jacobian matrix as a function of \mathbf{x} .
- (C) Is the Jacobian a full rank matrix for all of $\mathbf{x} \in \mathbb{R}^3$? Explain your reasoning.

Problem 5 (20pts)

Compute the gradients for the following expressions. (You can use identities, but show your work.)

(A) $\nabla_x \text{trace}(\mathbf{x}\mathbf{x}^T + \sigma^2 \mathbf{I})$ Assume $\mathbf{x} \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}$.

(B) $\nabla_x \frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ Assume $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^n$ and invertible symmetric $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$.

(C) $\nabla_x (\mathbf{c} - \mathbf{A}\mathbf{x})^T (\mathbf{c} - \mathbf{A}\mathbf{x})$ Assume $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$.

(D) $\nabla_x (\mathbf{c} + \mathbf{A}\mathbf{x})^T (\mathbf{c} - \mathbf{B}\mathbf{x})$ Assume $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{c} \in \mathbb{R}^m$ and $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$.

Problem 6 (2pts)

Approximately how many hours did this assignment take you to complete?

My notebook URL: <https://colab.research.google.com/XXXXXXXXXXXXXXXXXXXXXXXXX>

Changelog

- 23 November 2023 – Initial F23 version