Nicholas Carolan
email@princeton.edu
COS302-F23

Assignment #9

Due: 6:00pm Wednesday 29 November 2023

Discussants: N/A

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

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).

Use the first derivative to find stationary points:

$$f'(x) = 3x^2 + 12x - 3$$

$$0 = 3x^2 + 12x - 3$$
$$0 = 3(x^2 + 4x - 1)$$
$$0 = x^2 + 4x - 1$$
$$1 = x^2 + 4x$$
$$5 = x^2 + 4x + 4 = (x + 2)^2$$
$$\pm\sqrt{5} = x + 2$$
$$x = -2 \pm \sqrt{5}$$

Then use the second derivative to determine the kind of points.

$$f''(x) = 6x + 12$$

With $x = -2 - \sqrt{5}$: $f''(x) = 6(-2 - \sqrt{5}) + 12 = -6\sqrt{5} < 0 \quad \longrightarrow \quad -2 - \sqrt{5}$ is a local maximum.

With $x = -2 + \sqrt{5}$: $f''(x) = 6(-2 + \sqrt{5}) + 12 = 6\sqrt{5} > 0 \quad \longrightarrow \quad -2 + \sqrt{5}$ is a local minimum.

Problem 2 (20pts)

Consider the following scalar-valued function

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

(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 $\boldsymbol{\theta} = [3, \pi/2, 0]^T$.

(A)

$\frac{\partial f(x,y,z)}{\partial x} = 2xy$

$\frac{\partial f(x,y,z)}{\partial y} = x^2 + 6\cos(z + 6y)$

$\frac{\partial f(x,y,z)}{\partial z} = \cos(z + 6y)$

(B)

The gradient $\nabla f(x, y, z) = \begin{bmatrix} 2xy \\ x^2 + 6\cos(z + 6y) \\ \cos(z + 6y) \end{bmatrix}$

Evaluated at $\boldsymbol{\theta} = [3, \pi/2, 0]^T \longrightarrow \nabla f(3, \pi/2, 0) = \begin{bmatrix} 2 * 3 * \pi/2 \\ 3^2 + 6\cos(0 + 6(\pi/2)) \\ \cos(0 + 6(\pi/2)) \end{bmatrix} = \begin{bmatrix} 3\pi \\ 9 + 6 * -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 3\pi \\ 3 \\ -1 \end{bmatrix}$

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 $f(x_1, x_2, \ldots, x_n)$ 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.

(A)

$\frac{\partial}{\partial x}\frac{\partial}{\partial y} f(x, y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} x \sin(xy) = \frac{\partial}{\partial x} x^2 \cos(xy) = 2x \cos(xy) - x^2 y \sin(xy)$

(B)

$\frac{\partial}{\partial y}\frac{\partial}{\partial x} f(x, y) = \frac{\partial}{\partial y}\frac{\partial}{\partial x} x \sin(xy) = \frac{\partial}{\partial y} \sin(xy) + xy \cos(xy) = x \cos(xy) + x \cos(xy) - x^2 y sin(xy) = 2x \cos(xy) - x^2 y \sin(xy)$

Problem 4 (20pts)

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

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

(A) What is the Jacobian matrix of $f(\boldsymbol{x})$?

(B) Write the determinant of this Jacobian matrix as a function of $\boldsymbol{x}$.

(C) Is the Jacobian a full rank matrix for all of $\boldsymbol{x} \in \mathbb{R}^3$? Explain your reasoning.

(A)

$$J_f(x) = \begin{bmatrix} \partial_{x_1} \sin(x_1 x_2 x_3) & \partial_{x_2} \sin(x_1 x_2 x_3) & \partial_{x_3} \sin(x_1 x_2 x_3) \\ 0 & \partial_{x_2} \cos(x_2 + x_3) & \partial_{x_3} \cos(x_2 + x_3) \\ 0 & 0 & \partial_{x_3} e^{-x_3^2/2} \end{bmatrix} = \begin{bmatrix} x_2 x_3 \cos(x_1 x_2 x_3) & x_1 x_3 \cos(x_1 x_2 x_3) & x_1 x_2 \cos(x_1 x_2 x_3) \\ 0 & -\sin(x_2 + x_3) & -\sin(x_2 + x_3) \\ 0 & 0 & -x_3 e^{-x_3^2/2} \end{bmatrix}.$$

(B)

Because the Jacobian is upper triangular we can easily compute its determinant:

$$\det J_f(x) = (x_2 x_3 \cos(x_1 x_2 x_3))(-\sin(x_2 + x_3))(-x_3 e^{-x_3^2/2}) = x_2 x_3^2 \cos(x_1 x_2 x_3) \sin(x_2 + x_3) e^{-x_3^2/2}$$

(C)

No. The Jacobian is full rank (rank 3) if and only if the determinant does not equal zero. Consider obvious counterexamples, such as when $x_2 = 0$ or $x_3 = 0$.

Problem 5 (20pts)

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

(A) $\nabla_{\boldsymbol{x}} \operatorname{trace}(\boldsymbol{x}\boldsymbol{x}^T + \sigma^2 \boldsymbol{I})$ Assume $\boldsymbol{x} \in \mathbb{R}^n$ and $\sigma \in \mathbb{R}$.

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

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

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

(A)

$\nabla_{\boldsymbol{x}} \operatorname{trace}(\boldsymbol{x}\boldsymbol{x}^T + \sigma^2 \boldsymbol{I}) = \nabla_{\boldsymbol{x}} \operatorname{trace}(xx^T) = \nabla_{\boldsymbol{x}} x^T x = 2x$

(B)

$\nabla_{\boldsymbol{x}} \frac{1}{2}(\boldsymbol{x}-\mu)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mu) =$

(C)

$\nabla_{\boldsymbol{x}}(\boldsymbol{c}-\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{c}-\boldsymbol{A}\boldsymbol{x}) = \nabla_{\boldsymbol{x}}\ (c^T c - 2x^T A^T c + x^T A^T A x) = -2A^T c + 2A^T A x = 2A^T(Ax - c)$

(D)

$\nabla_{\boldsymbol{x}}(\boldsymbol{c}+\boldsymbol{A}\boldsymbol{x})^T(\boldsymbol{c}-\boldsymbol{B}\boldsymbol{x}) = \nabla_{\boldsymbol{x}}\ (c^T c - c^T B x + x^T A^T c - x^T A^T B x)$
$= -B^T c + A^T c - (A^T B + B^T A)x$ (Using identity $\nabla_x x^T M x = (M + M^T)x$)
$= (A^T - B^T)c - (A^T B - B^T A)x$

Problem 6 (2pts)

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

My notebook URL: N/A

Changelog

- 23 November 2023 – Initial F23 version