Department of Computer Science

Ramakrishna Mission Vivekananda Educational Research Institute, Belur Math

Optimization for ML –BDA 2023

Problem Set on Matrix Calculus

1. Recall the definition of the derivative of a scalar with respect to a matrix ($dy/d\mathbf{X}$). We will now check if this is a valid extension of the scalar and vector case. Evaluate the derivatives when $\mathbf{X} \in \mathbb{R}^{1\times 1}$, $\mathbf{X} \in \mathbb{R}^{n\times 1}$, and $\mathbf{X} \in \mathbb{R}^{1\times n}$. Which definition does each of them correspond to?

2. We have derived that $d(\mathbf{A}\mathbf{x})/d\mathbf{x} = \mathbf{A}^T$ for $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{A} \in \mathbb{R}^{n\times p}$ that does not depend on $\mathbf{x}$. $\mathbf{A}\mathbf{x}$ results in a vector, and thus we have used the $d\mathbf{y}/d\mathbf{x}$ definition. Now consider $d(\mathbf{x}^T\mathbf{B})/d\mathbf{x}$ for $\mathbf{B} \in \mathbb{R}^{p\times n}$. Recall that the definition of $d\mathbf{y}/d\mathbf{x}$ does not change even when $\mathbf{y}$ is a row vector. Evaluate $d(\mathbf{x}^T\mathbf{B})/d\mathbf{x}$.

3. The quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is a form we will encounter often.* In this question, we are interested in $d(\mathbf{x}^T\mathbf{A}\mathbf{x})/d\mathbf{x}$. Assume that $\mathbf{A}$ is not a function of $\mathbf{x}$.

   (a) Evaluate $\mathbf{x}^T\mathbf{A}\mathbf{x}$ when $\mathbf{x} = [x_1, x_2]^T$ and the (i, j)-th element of $\mathbf{A}$ is A_{ij}. Why do you think $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is called the quadratic form?

   (b) Which definition of the derivative do we need in order to evaluate $d(\mathbf{x}^T\mathbf{A}\mathbf{x})/d\mathbf{x}$?

   (c) Assume $\mathbf{x} \in \mathbb{R}^2$ and $\mathbf{A} \in \mathbb{R}^{2\times 2}$. Evaluate $d(\mathbf{x}^T\mathbf{A}\mathbf{x})/d\mathbf{x}$.

   (d) Generalize the previous result to when $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n\times n}$ and evaluate $d(\mathbf{x}^T\mathbf{A}\mathbf{x})/d\mathbf{x}$. Can you express the result in matrix form?

   (e) What happens when $\mathbf{A}$ is a symmetric matrix, i.e., $\mathbf{A}^T = \mathbf{A}$?

4.

Evaluate $\partial f/\partial x$ and $\partial f/\partial y$ for each of the following:*

   (a) $f(u, v) = (u - v)e^u$, where $u = xy$ and $v = x^2 - y^2$

   (b) $f(u, v) = u \log v + v \log u$, where $u = \frac{x}{2} + \frac{2}{y}$ and $v = xe^y$

   (c) $f(u, v) = u \log v$, where $u = x \sin y + y \sin x$ and $v = x \cos y + y \cos x$

   (d) $f(u, v) = (u + v)/(1 - uv)$, where $u = \tan \frac{x+y}{2}$ and $v = \tan \frac{x-y}{2}$

5.

   (a) Evaluate $\frac{d}{dx}\sigma(x)$ where $\sigma(x) = 1/(1 + e^{-x})$. This is called the sigmoid function.

   (b) Express your answer in (a) using only $\sigma(x)$ and constants.

   (c) Evaluate $\frac{d}{dx}\tanh(x)$ where $\tanh(x) = (e^x - e^{-x})/(e^x + e^{-x})$. This is called the hyperbolic tangent function.

   (d) Express your answer in (c) using only $\tanh(x)$ and constants.

6.

(a) $\nabla f(x, y)$ where $f(x, y) = xy^2 + x^2y$

(b) $\nabla f(x, y)$ where $f(x, y) = (x + y)^2$

(c) $\nabla^2 f(x, y)$ where $f(x, y) = \sin(e^{xy})$

(d) $\nabla f(\mathbf{x})$ where $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$

(e) Express your answer in (d) using only one variable (no limit on constants).

(f) $\nabla f(\mathbf{x})$ where $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ and $\mathbf{w}$ is a constant vector

(g) Express your answer in (f) using only one variable (no limit on constants).

7.

(a) $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ where $f(x, y) = x^y + y^x$

(b) $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ where $f(x, y) = \sin(y + \cos x)$

(c) $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ where $f(x, y) = e^{xy} + y \log 3x$

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

(e) $\nabla_x f(x, y)$ and $\nabla_y f(x, y)$ where $f(x, y) = x^{\log y} + x^2 + 2y$

(f) $\nabla_x f(x, y)$ and $\nabla_y f(x, y)$ where $f(x, y) = (x + y)^2$

(g) $\frac{\partial f}{\partial x_i}$ where $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$ $(1 \leq i \leq n)$ *Hint:* Recall that $\|\mathbf{x}\|_2 = \sqrt{x_1^2 + \cdots + x_n^2}$

(h) $\frac{\partial f}{\partial x_i}$ where $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ and $\mathbf{w}$ is a constant vector $(1 \leq i \leq n)$