Theoretical Machine Learning Theoretical Assignment 1

Naman Mohan Singh

Problem 1

(a) Prove that if A is an $n \times m$ matrix and $\mathbf{x}$ is an $m \times 1$ vector, then:

$$d[A\mathbf{x}] = A\,d\mathbf{x}$$

Proof:
Let $\mathbf{x}(t)$ be a function of t such that $\mathbf{x}(t) \in \mathbb{R}^m$. Then,

$$\mathbf{y}(t) = A\mathbf{x}(t)$$

where $\mathbf{y}(t) \in \mathbb{R}^n$.
The differential of $\mathbf{y}$ with respect to t is given by:

$$d\mathbf{y} = \frac{d\mathbf{y}}{dt}dt = \frac{d}{dt}(A\mathbf{x}(t))dt = A\frac{d\mathbf{x}}{dt}dt$$

Therefore,

$$d[A\mathbf{x}] = A\,d\mathbf{x}$$

(b) Prove that if A is an $n \times n$ matrix and $\mathbf{x}$ is an $n \times 1$ vector, then:

$$d[\mathbf{x}^T A\mathbf{x}] = \mathbf{x}^T(A + A^T)d\mathbf{x}$$

Proof:
Let $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. The differential of f is given by:

$$d[f(\mathbf{x})] = d[\mathbf{x}^T A\mathbf{x}]$$

Using the product rule for differentials,

$$d[\mathbf{x}^T A\mathbf{x}] = (d\mathbf{x})^T A\mathbf{x} + \mathbf{x}^T A\,d\mathbf{x}$$

Since $(d\mathbf{x})^T = (d\mathbf{x})^T$ and A is a matrix,

$$(d\mathbf{x})^T A\mathbf{x} + \mathbf{x}^T A\,d\mathbf{x} = \mathbf{x}^T A^T d\mathbf{x} + \mathbf{x}^T A\,d\mathbf{x}$$

$$d[\mathbf{x}^T A\mathbf{x}] = \mathbf{x}^T(A^T + A)d\mathbf{x}$$

Problem 2

Suppose you have a matrix of dimension $m \times n$, and you are differentiating with respect to a $k \times 1$ vector. What is the dimension of the final result?

Solution:

Let A be an $m \times n$ matrix and $\mathbf{y}$ be a $k \times 1$ vector. If we are differentiating A with respect to $\mathbf{y}$, the result will be a tensor with dimensions corresponding to the dimensions of A and $\mathbf{y}$.

The resulting dimension will be $m \times n \times k$.

Problem 3

Solve:

(a)

$$\frac{d}{d\begin{bmatrix} x \\ y \end{bmatrix}} \left(2\sin^2(x)\cos(y)\right)$$

Solution:

Let $f(x, y) = 2\sin^2(x)\cos(y)$.

$$\frac{\partial f}{\partial x} = 4\sin(x)\cos(x)\cos(y)$$

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

So,

$$\frac{d}{d\begin{bmatrix} x \\ y \end{bmatrix}} \left(2\sin^2(x)\cos(y)\right) = \begin{bmatrix} 4\sin(x)\cos(x)\cos(y) \\ -2\sin^2(x)\sin(y) \end{bmatrix}$$

(b)

$$\frac{d}{d\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}} \begin{bmatrix} 3x^2y + xyzw \\ \sin(x^2 + yw - z) \end{bmatrix}$$

Solution:

Let $g_1(x, y, z, w) = 3x^2y + xyzw$ and $g_2(x, y, z, w) = \sin(x^2 + yw - z)$.

$$\frac{\partial g_1}{\partial x} = 6xy + yzw$$

$$\frac{\partial g_1}{\partial y} = 3x^2 + xzw$$

$$\frac{\partial g_1}{\partial z} = xyw$$

$$\frac{\partial g_1}{\partial w} = xyz$$

$$\frac{\partial g_2}{\partial x} = \cos(x^2 + yw - z) \cdot 2x$$

$$\frac{\partial g_2}{\partial y} = \cos(x^2 + yw - z) \cdot w$$

$$\frac{\partial g_2}{\partial z} = -\cos(x^2 + yw - z)$$

$$\frac{\partial g_2}{\partial w} = \cos(x^2 + yw - z) \cdot y$$

So,

$$\frac{d}{d\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}} \begin{bmatrix} 3x^2y + xyzw \\ \sin(x^2 + yw - z) \end{bmatrix} = \begin{bmatrix} 6xy + yzw & 3x^2 + xzw & xyw & xyz \\ 2x\cos(x^2 + yw - z) & w\cos(x^2 + yw - z) & -\cos(x^2 + yw - z) & y\cos(x^2 + \end{bmatrix}$$

Problem 4 (Bonus)

Solve for:

$$\frac{d}{d\mathbf{x}} \left(e^{\beta^T \mathbf{x}} \right)$$

where β is an $n \times 1$ vector and $\mathbf{x}$ is an $n \times 1$ vector.
Solution:
Let $f(\mathbf{x}) = e^{\beta^T \mathbf{x}}$.
Using the chain rule,

$$\frac{d}{d\mathbf{x}} \left(e^{\beta^T \mathbf{x}} \right) = e^{\beta^T \mathbf{x}} \frac{d}{d\mathbf{x}} (\beta^T \mathbf{x})$$

Since $\frac{d}{d\mathbf{x}}(\beta^T \mathbf{x}) = \beta$,

$$\frac{d}{d\mathbf{x}} \left(e^{\beta^T \mathbf{x}} \right) = e^{\beta^T \mathbf{x}} \beta$$

Using Question 1. In above problem