# Theoretical Machine Learning
(Theoretical Assignment 1)

Akshat Sharma

May 13, 2024

Problem 1.

(a). Let A= $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$ and x= $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Therefore, their product A.x= $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n a_{1i}x_i \\ \sum_{i=1}^n a_{2i}x_i \\ \vdots \\ \sum_{i=1}^n a_{mi}x_i \end{pmatrix}$

The derivative $\frac{d(A.x)}{dx} = \begin{pmatrix} \frac{\partial \sum_{i=1}^n a_{1i}x_i}{\partial x_1} & \frac{\partial \sum_{i=1}^n a_{1i}x_i}{\partial x_2} & \dots & \frac{\partial \sum_{i=1}^n a_{1i}x_i}{\partial x_n} \\ \frac{\partial \sum_{i=1}^n a_{2i}x_i}{\partial x_1} & \frac{\partial \sum_{i=1}^n a_{2i}x_i}{\partial x_2} & \dots & \frac{\partial \sum_{i=1}^n a_{2i}x_i}{\partial x_n} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial \sum_{i=1}^n a_{mi}x_i}{\partial x_1} & \frac{\partial \sum_{i=1}^n a_{mi}x_i}{\partial x_2} & \dots & \frac{\partial \sum_{i=1}^n a_{mi}x_i}{\partial x_n} \end{pmatrix}$.

$= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$=A (b). Let A= $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$ and x= $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Then $x^T Ax = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$

$= \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} \sum_{i=1}^n a_{1i}x_i \\ \sum_{i=1}^n a_{2i}x_i \\ \vdots \\ \sum_{i=1}^n a_{ni}x_i \end{pmatrix}$

$= x_1 \sum_{i=1}^n a_{1i}x_i + x_2 \sum_{i=1}^n a_{2i}x_i + \dots + x_n \sum_{i=1}^n a_{ni}x_i$, which is a scalar.

Let p=$x_1 \sum_{i=1}^n a_{1i}x_i + x_2 \sum_{i=1}^n a_{2i}x_i + \dots + x_n \sum_{i=1}^n a_{ni}x_i$.

The derivative $\frac{d(x^T Ax)}{dx}$= $\begin{pmatrix} \frac{\partial(x^T Ax)}{\partial x_1} & \frac{\partial(x^T Ax)}{\partial x_2} & \dots & \frac{\partial(x^T Ax)}{\partial x_n} \end{pmatrix}$

$= \begin{pmatrix} \sum_{i=1}^n x_i(a_{i1} + a_{1i}) & \sum_{i=1}^n x_i(a_{i2} + a_{2i}) & \dots & \sum_{i=1}^n x_i(a_{in} + a_{ni}) \end{pmatrix}$

Now $A + A^T = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & \dots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix}$

$= \begin{pmatrix} a_{11} + a_{11} & a_{12} + a_{21} & \dots & a_{1n} + a_{n1} \\ a_{21} + a_{12} & a_{22} + a_{22} & \dots & a_{2n} + a_{n2} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1} + a_{1n} & a_{n2} + a_{2n} & \dots & a_{nn} + a_{nn} \end{pmatrix}$

and $x^T = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}$.

Therefore, $x^T(A{+}A^T){=}\begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} a_{11}+a_{11} & a_{12}+a_{21} & \dots & a_{1n}+a_{n1} \\ a_{21}+a_{12} & a_{22}+a_{22} & \dots & a_{2n}+a_{n2} \\ \vdots & \vdots & \dots & \vdots \\ a_{n1}+a_{1n} & a_{n2}+a_{2n} & \dots & a_{nn}+a_{nn} \end{pmatrix}$

$= \begin{pmatrix} \sum_{i=1}^{n} x_i(a_{i1}+a_{1i}) & \sum_{i=1}^{n} x_i(a_{i2}+a_{2i}) & \dots & \sum_{i=1}^{n} x_i(a_{in}+a_{ni}) \end{pmatrix}.$

Thus we see that $\dfrac{d(x^TAx)}{dx} = x^T(A+A^T)$.

Problem 2. Consider a matrix of dimension m x n

A= $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$

and a k x 1 vector x= $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{pmatrix}$.

Differentiating a matrix with respect to a vector is equivalent to differentiating each column of the matrix, treated as an individual vector, with respect to the vector x, and then placing the obtained matrices row-wise to obtain the final derivative matrix.
Each column of the matrix A can be treated as a (m x 1) vector. The derivative of each column w.r.t. x, therefore, is a matrix of dimension (m x k). Since there are n columns in the original matrix, we obtain n such (m x k) matrices. Stacking them horizontally yields a matrix of dimension (m x (n x k)).
Therefore, the final result is of dimension (m x (n x k)).

Problem 3. (a). We have, $v_1{=}\begin{pmatrix} 2sin^2(x)cos(y) \\ x^2+3e^y \end{pmatrix}$ and $v_2{=}\begin{pmatrix} x \\ y \end{pmatrix}$

Now $\dfrac{dv_1}{dv_2} = \begin{pmatrix} \dfrac{\partial(2sin^2(x)cos(y))}{\partial x} & \dfrac{\partial(2sin^2(x)cos(y))}{\partial y} \\ \dfrac{\partial(x^2+3e^y)}{\partial x} & \dfrac{\partial(x^2+3e^y)}{\partial y} \end{pmatrix} = \begin{pmatrix} 4sin(x)cos(x)cos(y) & -2sin^2(x)sin(y) \\ 2x & 3e^y \end{pmatrix}$.

(b). Here $v_1 = \begin{pmatrix} 3x^2y + xyzw \\ sin(x^2 + yw - z) \end{pmatrix}$ and $v_2 = \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix}$

$$\frac{dv_1}{dv_2} = \begin{pmatrix} \dfrac{\partial(3x^2y+xyzw)}{\partial x} & \dfrac{\partial(3x^2y+xyzw)}{\partial y} & \dfrac{\partial(3x^2y+xyzw)}{\partial z} & \dfrac{\partial(3x^2y+xyzw)}{\partial w} \\ \dfrac{\partial(sin(x^2+yw-z))}{\partial x} & \dfrac{\partial(sin(x^2+yw-z))}{\partial y} & \dfrac{\partial(sin(x^2+yw-z))}{\partial z} & \dfrac{\partial(sin(x^2+yw-z))}{\partial w} \end{pmatrix}$$

$$= \begin{pmatrix} 6xy + yzw & 3x^2 + xzw & xyw & xyz \\ 2xcos(x^2+yw-z) & wcos(x^2+yw-z) & -cos(x^2+yw-z) & ycos(x^2+yw-z). \end{pmatrix}$$