Matrix Calculus

Daniel Yu

September 18, 2024

1 Notational Rules

In standard linear algebra notation $\vec{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ a column vector. Thus, a row vector is the transpose $\vec{x}^T = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$

However, it is machine learning convention to write the matrices as:

$$X = \begin{pmatrix} \vec{x}_1^T \\ \vdots \\ \vec{x}_n^T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ \vdots & & & \\ x_{n_1} & x_{n_2} & \dots & x_{nm} \end{pmatrix}.$$

with the row vectors because in this case row vectors can be thought of the inputs $\vec{x}^i$ for a datapoint $(\vec{x}^i, y^i)$, so X is the entire training set of input vectors $\vec{x}^i$ with each $\vec{x}^i$ representing one set of inputs.

2 "Derivative" Matrix Rules

Note. Ax
Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, so $Ax = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{pmatrix}$ with $f_1 = x_1 + 2x_2$ and $f_2 = 3x_1 + 4x_2$.

$$\frac{d}{d\vec{x}} A\vec{x} = \frac{d}{d\vec{x}} \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{pmatrix} = \begin{pmatrix} \frac{df_1}{x_1} & \frac{df_1}{x_2} \\ \frac{df_2}{x_1} & \frac{df_2}{x_2} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = A.$$

Remark. Note that we are not taking the derivative of A since A is a constant matrix. We are taking the derivative of $A\vec{x}$, i.e. taking derivative of A as a linear transformation upon some vector of variables $\vec{x}$!!!!

Definition 1.

$$\frac{d}{d\vec{x}} A\vec{x} = A.$$

Note. $x^T Ax$
Let $\vec{x}^T A\vec{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = a_{11}x_1^2 + a_{12}x_1x_2 + a_{21}x_1x_2 + a_{22}x_2^2 = f(x_1, x_2)$, so

$$\begin{aligned} \frac{d}{d\vec{x}}[\vec{x}^T A\vec{x}] &= \frac{d}{d\vec{x}} f(x_1, x_2) \\ &= \begin{pmatrix} \frac{df}{dx_1} \\ \frac{df}{dx_2} \end{pmatrix} \\ &= \begin{pmatrix} 2a_{11}x_1 + a_{12}x_2 + a_{21}x_2 \\ a_{12}x_1 + a_{21}x_1 + 2a_{22}x_2 \end{pmatrix} \\ &= \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix} + \begin{pmatrix} a_{11}x_1 + a_{21}x_2 \\ a_{12}x_1 + a_{22}x_2 \end{pmatrix} \\ &= (A + A^T)\vec{x}. \end{aligned}$$

Definition 2.

$$\frac{d}{d\vec{x}} \vec{x}^T A\vec{x} = (A + A^T)\vec{x}.$$

Corollary 1. When A is symmetric, $A = A^T$:

$$\frac{d}{d\vec{x}} \vec{x}^T A\vec{x} = 2A\vec{x}.$$