Matrix & vector differentiation

Let $y = f(x)$

$y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$

both are vectors

The Jacobian matrix is

$$J = \frac{\partial y}{\partial x} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & & & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

if x is a scalar, then the resulting Jacobian $J \in \mathbb{R}^{m \times 1}$

Define

$y \in Ax$

$y \in \mathbb{R}^m \quad x \in \mathbb{R}^n$

$A \in \mathbb{R}^{m \times n}$

$$\frac{\partial y}{\partial x} = A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ \vdots & & & \\ a_{m1} & & \cdots & a_{mn} \end{bmatrix}$$

To see this write ith element of y as

$$y_i = \sum_{k=1}^{n} a_{ik} x_k$$

$$\frac{\partial y_i}{\partial x_j} = a_{ij} \quad \text{for all } i = 1, 2, \ldots m$$
$$j = 1, 2, \ldots n$$

$$\Rightarrow \quad \frac{\partial y}{\partial x} = A$$

Define

$y = Ax$ as before with respect to dim.

A does not depend on x

x depends on z (vector)

$$\frac{\partial y}{\partial z} = A \frac{\partial x}{\partial z}$$

$$y_i = \sum_{k=1}^{n} a_{ik} x_k$$

$$\frac{\partial y_i}{\partial z_j} = \underbrace{\sum_{k=1}^{n} a_{ik} \frac{\partial x_k}{\partial z_j}}$$

element (i,j) of $A \dfrac{\partial x}{\partial z}$

Hence

$$\frac{\partial y}{\partial z} = \frac{\partial y}{\partial x} \frac{\partial x}{\partial z} = A \frac{\partial x}{\partial z}$$

Define a scalar α

$$\alpha = y^T A x$$

$y \in \mathbb{R}^m$ $x \in \mathbb{R}^n$

$A \in \mathbb{R}^{m \times n}$

A is independent of y and x

$$\frac{\partial \alpha}{\partial y} = x^T A^T \wedge \frac{\partial \alpha}{\partial x} = y^T A$$

Proof: Define $w^T = y^T A$

$\alpha = w^T x$

$$\frac{\partial \alpha}{\partial x} = w^T = y^T A$$

α is a scalar:

$$\alpha = \alpha^T = x^T A^T y$$

$$\frac{\partial \alpha}{\partial y} = x^T A^T$$

Define

$$\alpha = x^T A x$$

A is a quadratic matrix $A \in \mathbb{R}^{n \times n}$

$$\frac{\partial \alpha}{\partial x} = x^T (A + A^T)$$

Proof

$$\alpha = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_i x_j$$

$$\frac{\partial \alpha}{\partial x_k} = \sum_{j=1}^{n} a_{kj} x_j + \sum_{i=1}^{n} a_{ik} x_i$$

for all $k = 1, 2, \ldots n$, this leads to

$$\frac{\partial \alpha}{\partial x} = x^T A^T + x^T A$$

$$= x^T (A^T + A)$$

if A is symmetric

$$\frac{\partial \alpha}{\partial x} = 2 x^T A$$

Define

$$\alpha = y^T x = \text{scalar}$$

$$y, x \in \mathbb{R}^n$$

Both are function of a vector z

$$\frac{\partial \alpha}{\partial z} = x^T \frac{\partial y}{\partial z} + y^T \frac{\partial x}{\partial z}$$

$$\partial z \qquad\qquad \partial z \quad \cdot \quad \partial z$$

Proof

$$\alpha = \sum_{j=1}^{n} x_i y_i$$

$$\frac{\partial \alpha}{\partial z_k} = \sum_{j=1}^{n} \left(x_i \frac{\partial y_i}{\partial z_k} + y_i \frac{\partial x_i}{\partial z_k} \right)$$

for all $k = 1, 2, \ldots n \Rightarrow$

$$\frac{\partial \alpha}{\partial z} = \frac{\partial \alpha}{\partial y}\frac{\partial y}{\partial z} + \frac{\partial \alpha}{\partial x}\frac{\partial x}{\partial z}$$

$$= x^T \frac{\partial y}{\partial z} + y^T \frac{\partial x}{\partial z}$$

Define

$$x^T x = \alpha \quad \text{(scalar)}$$

$$\frac{\partial \alpha}{\partial z} = 2x^T \frac{\partial x}{\partial z}$$

Define scalar α

$$\alpha = y^T A x$$

no dependence on z

Depend on z

$$\frac{\partial \alpha}{\partial z} = x^T A^T \frac{\partial y}{\partial z} + y^T A \frac{\partial x}{\partial z}$$

Proof $\quad w^T = y^T A$

$$\alpha = w^T x$$

$$\frac{\partial \alpha}{\partial z} = x^T \frac{\partial w}{\partial z} + w^T \frac{\partial x}{\partial z}$$

substitute back w

$$\frac{\partial \alpha}{\partial z} = \frac{\partial \alpha}{\partial y}\frac{\partial y}{\partial z} + \frac{\partial \alpha}{\partial x}\frac{\partial x}{\partial z}$$

$$= x^T A^T \frac{\partial y}{\partial z} + y^T A \frac{\partial x}{\partial z}$$

Finaly :

$$\alpha = (y - Ax)^T (y - Ax)$$

(used in MSE definition)

Define $w = y - Ax$

w depends on X

$$\alpha = w^T w$$

$$\frac{\partial \alpha}{\partial x} = 2 w^T \frac{\partial w}{\partial x}$$

$$\frac{\partial w}{\partial x} = -A \implies$$

$$\frac{\partial \alpha}{\partial x} = -2(y - Ax)^T A$$

$$= -2y^T A + 2x^T A^T A$$

does not depend on X

$$\frac{\partial^2 \alpha}{\partial x^T \partial x} = 2A^T A$$