MA19 (E)

$f: \mathbb{R}^n \longrightarrow \mathbb{R}$

$f: \mathbb{R}^{n\times m} \longrightarrow \mathbb{R}$ $\qquad f(A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}$

~~Mat~~

$f: \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$ $\qquad \det(A)$ $\qquad \det: \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}$

$\text{trace}(A) = \sum_{i=1}^{n} a_{ii}$

$$\nabla f: \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}} \\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{a_{22}} & \cdots & \frac{\partial f}{\partial a_{2n}} \\ \vdots & & & \vdots \\ \frac{\partial f}{\partial a_{m1}} & & & \frac{\partial f}{\partial a_{mn}} \end{bmatrix} \qquad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \vdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$1_n = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \in \mathbb{R}^n$$

$$f(A) = \sum_i \sum_j a_{ij} = 1_m^T A 1_n$$

$$\frac{\partial f}{\partial a_{kl}} = 1 \qquad \nabla f = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} = 1_m 1_n^T$$

$$\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}}_{1\times n}$$

$$D[U]f = \frac{d}{d\alpha} f(\underset{\uparrow\,\mathbb{R}^{m\times n}}{A} + \alpha \underset{\uparrow\,\mathbb{R}^{m\times n}}{U}) = \frac{d}{d\alpha} 1_m^T (A + \alpha V) \overset{m\times 1}{1_n} = \frac{d}{d\alpha} 1_m^T A 1_n + \alpha 1_m^T U 1_n$$

$$= 1_m^T U 1_n = \langle 1_m, U 1_n \rangle = \langle 1_m 1_n^T, U \rangle$$

$$\Rightarrow \nabla f = 1_m 1_n^T = 1_{m\times n}$$

MA 19 II

$$\text{trace}: \mathbb{R}^{n\times n} \longrightarrow \mathbb{R} \qquad \text{trace}(A) = \sum_{i=1}^{n} a_{ii}$$

$$\frac{\partial\, \text{trace}}{a_{ij}} = \begin{cases} 0 & i = j \\ 1 & i = j \end{cases}$$

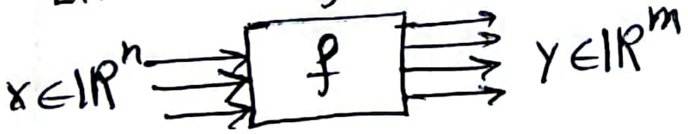
$$\nabla \text{trace} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = I_n$$

$$A \in \mathbb{R}^{m\times n} \qquad u \in \mathbb{R}^m \quad v \in \mathbb{R}^n$$

$$f(A) = u^T A v \qquad \nabla f = u v^T \in \mathbb{R}^{m\times n}$$

$$= \sum_i \sum_j u_i v_j a_{ij}$$

Linear Regression



$$f_\theta(x) = f(\theta, x) = A\vec{x} + \vec{b}$$

$$\theta = (A, \vec{b}) \quad A \in \mathbb{R}^{m\times n}$$

$$b \in \mathbb{R}^m$$

$$x_1, x_2, \dots, x_N \in \mathbb{R}^n \qquad (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$$

$$y_1 \quad y_2 \qquad y_N \in \mathbb{R}^m$$

$$C(\theta) = C(A,b) = \sum_{i=1}^{N} \| A x_i + b - y_i \|^2 = \sum_{i=1}^{N} (A x_i + b - y_i)^T (A x_i + b - y_i)$$

$$\cancel{\nabla C} \quad \nabla_b C = \frac{\partial C}{\partial b}$$

$$D[u]\, C(A, \cancel{b+\alpha u}) = \frac{d}{d\alpha} C(A, b+\alpha u) \Big|_{\alpha=0}$$

$$\frac{d}{d\alpha} \sum_{i=1}^{N} (A x_i + b + \alpha u - y_i)^T (A x_i + b + \alpha u - y_i) \Big|_{\alpha=0} = \sum_{i=1}^{N} 2 u^T (A x_i + b - y_i)$$

$$= u^T \left(2 \sum_{i=1}^{N} (A x_i + b - y_i) \right) \Rightarrow \nabla_b$$

MA19 (III)

$$\nabla_b C(A,b) = 2\sum_{i=1}^{N}(Ax_i + b - y_i)$$

$$\nabla_b = 0 \Rightarrow \sum_{i=1}^{N}(Ax_i + b - y_i) = 0 \Rightarrow A\sum_{i=1}^{N} x_i + n\vec{b} - \sum_{i=1}^{N} y_i = 0$$

$$\vec{b} = \frac{1}{n}\sum_{i=1}^{N} y_i - A\frac{1}{n}\sum_{i=1}^{N} x_i$$

$$\vec{b} = \bar{y} - A\bar{x}$$

$$C(A,b) = \sum_{i=1}^{N}\|Ax_i + b - y_i\|^2 = \sum_{i=1}^{N}\|Ax_i + \bar{y} - A\bar{x} - y_i\|^2$$

$$= \sum_{i=1}^{N}\|A\underbrace{(x_i - \bar{x})}_{\tilde{x}_i} - \underbrace{(y_i - \bar{y})}_{\tilde{y}_i}\|^2$$

$$C(A) = \sum_{i=1}^{N}\|Ax_i - y_i\|^2 = \sum_{i=1}^{N}(Ax_i - y_i)^T(Ax_i - y_i)$$

$$\frac{\partial C(A)}{\partial A} = ? \qquad D[U]C(A) = \frac{d}{d\alpha}C(A + \alpha U)\Big|_{\alpha=0}$$

$$\frac{d}{d\alpha}\sum_{i=1}^{N}\left((A + \alpha U)x_i - y_i\right)^T\left((A + \alpha U)x_i - y_i\right)\Big|_{\alpha=0}$$

$$= 2\sum_{i=1}^{N}(Ux_i)^T(Ax_i - y_i) = 2\sum_{i=1}^{n}\langle Ux_i, Ax_i - y_i\rangle$$

$$= 2\sum_{i=1}^{N}\langle U, (Ax_i - y_i)x_i^T\rangle = \langle U, 2\sum_{i=1}^{n}(Ax_i - y_i)x_i^T\rangle$$

$$\nabla_A C = 2\sum_{i=1}^{N}(Ax_i - y_i)x_i^T$$

MA19 IV

$$\nabla_A C = 0 \Rightarrow \sum_{i=1}^{N} \underbrace{(\underbrace{A}_{m\times n} \underbrace{x_i}_{n\times 1} - \underbrace{y_i}_{m\times 1})}_{m\times 1} \underbrace{x_i^T}_{1\times n} = 0$$

$m \times n$

$$\Rightarrow \sum_{i=1}^{N} A x_i x_i^T - y_i x_i^T = 0$$

$$\Rightarrow A \sum_{i=1}^{N} x_i x_i^T = \sum_{i=1}^{N} y_i x_i^T$$

$$\Rightarrow A = \underbrace{\left(\sum_{i=1}^{N} y_i x_i^T\right)}_{} \underbrace{\left(\sum_{i=1}^{N} x_i x_i^T\right)}_{}^{-1}$$

$$X = [x_1 - x_N] \qquad (YX^T)\,(\underline{XX^T})^{-1}$$

$$Y = [y_1 - y_N]$$

$$f: \mathbb{R}^n \longrightarrow \mathbb{R}^m \qquad \underset{x \in \mathbb{R}^n}{f(x)} = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix} \qquad f_i(x): \mathbb{R}^n \to \mathbb{R}$$

Linearization $\cancel{f(x_0)}$ $f(x) \simeq f(x_0) + \ell\,(x - x_0)$

x around x_0 — ℓ linear

$\ell: \mathbb{R}^n \to \mathbb{R}^m$

$$f(x) \simeq f(x_0) + \overbrace{M}^{m\times n}(x - x_0)$$

ماتریس یعقوبی

Jacobian $\quad M = M(x_0)$

$$f: \mathbb{R}^n \to \mathbb{R} \qquad f(x) = f(x_0) + \nabla^T (x - x_0)$$

Scanned with CamScanner

MA 19 ∇

$$f(x) = f(x_0) + M\,(x - x_0)$$

$$\begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix} = \begin{bmatrix} f_1(x_0) \\ f_2(x_0) \\ \vdots \\ f_m(x_0) \end{bmatrix} + \begin{bmatrix} m_1^T \\ m_2^T \\ | \\ m_n^T \end{bmatrix} (x - x_0) \qquad M = \begin{bmatrix} m_1^T \\ m_2^T \\ \vdots \\ m_n^T \end{bmatrix}$$

$$f_1(x) = f_1(x_0) + \vec{m}_1^T (x - x_0) = f_1(x_0) + \nabla f_1^T (x - x_0)$$

$$f_2(x) = f_2(x_0) + \vec{m}_2^T (x - x_0)$$

$$\vdots$$

$$f_m(x) : f_m(x_0) = \vec{m}_m^T (x - x_0)$$

$$M = J = \begin{bmatrix} \nabla f_1^T \\ \nabla f_2^T \\ \vdots \\ \nabla f_m^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \text{---} & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

$$f(x) \simeq f(x_0) + J_f(x_0)\,(x - x_0)$$

$\cancel{D[u]f = f(x + \alpha u)}$

$$D[u]f(x) = \frac{d}{d\alpha} f(x + \alpha u) = \boxed{J u}$$

$$f(x) = \underbrace{A}_{m \times n} x \qquad D[u]f(x) = \frac{d}{d\alpha} A(x + \alpha u)$$

$$= \underbrace{A}_{J} u$$

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

$f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$

$$f(x) = \underbrace{\underbrace{x}_{n\times 1}\ \underbrace{x^T}_{1\times n}\ \underbrace{A}_{n\times n}\ \underbrace{x}_{n\times 1}}_{n\times 1}$$

$$D[u]f(x) = \frac{d}{d\alpha} (x+\alpha u)(x+\alpha u)^T A (x+\alpha u) \Big|_{\alpha=0}$$

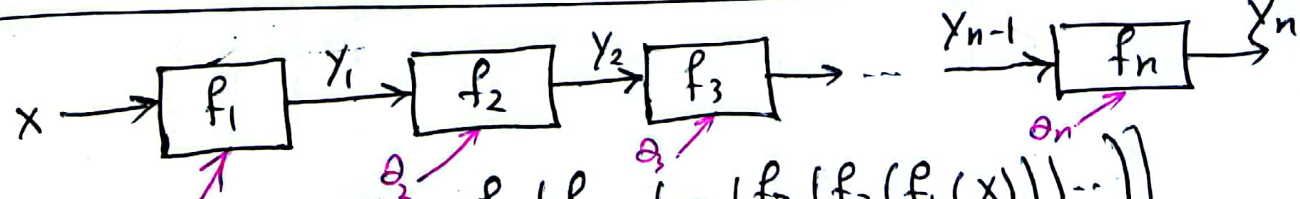
$$= u(x+\alpha u)^T A (x+\alpha u) + (x+\alpha u) u^T A (x+\alpha u) + (x+\alpha u)(x+\alpha u)^T A u \Big|_{\alpha=0}$$

$$= u \overbrace{x^T A x}^{1\times 1} + x \overbrace{u^T A x}^{1\times 1} + x x^T A u$$

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

$$\Big(\underbrace{(x^T A x) I + x x^T A^T + x x^T A}_{J} \Big) u$$

$$2u = \begin{bmatrix} 2 & & \\ & 2 & \\ & & \ddots \\ & & & 2 \end{bmatrix} u = (2I) u$$



$$Y_n = f_n(f_{n-1}(\cdots(f_3(f_2(f_1(x)))\cdots)))$$

$$Y_n = f_n(\theta_n, f_{n-1}(\theta_{n-1}, f_{n-2}(\cdots, f_2(\theta_2, f_1(\theta_1, x_1))\cdots)))$$

$$\theta = (\theta_n, \theta_{n-1}, \cdots, \theta_1)$$

$$f_i = \sigma(A^i Y_{i-1} + b^i)$$