Mathematics for AI

Lecture 17

functions of matrices, linear regression, vector valued multivariate functions, Jacobian

Functions on matrices

$$f: \mathbb{R}^{m\times n} \longrightarrow \mathbb{R} \qquad \nabla_f \in \mathbb{R}^{m\times n}$$

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

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

$$f(A) = f\left(\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix}\right) \qquad \nabla_f = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}} \\ \cdots & & & \\ \frac{\partial f}{\partial a_{m1}} & \frac{\partial f}{\partial a_{m2}} & \cdots & \frac{\partial f}{\partial a_{mn}} \end{bmatrix}$$

Functions on matrices, Example

$\cancel{f(A)}\quad f: \mathbb{R}^{n\times n} \rightarrow \mathbb{R} \qquad A \mapsto \cancel{\tan}\,\mathrm{trace}(A)$

$$f(A) = \mathrm{trace}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

$$\nabla_f = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Functions on matrices, Example 2

K. N. Toosi
University of Technology

$$f: \mathbb{R}^{m\times n} \longrightarrow \mathbb{R} \qquad f(A) = u^T A v \quad \text{linear w.r.t } A \qquad \left.\begin{array}{l} u \in \mathbb{R}^m \\ v \in \mathbb{R}^n \end{array}\right\} \text{constant}$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} u_i v_j$$

$U \in \mathbb{R}^{m\times n}$ direction

$$\nabla_f = \frac{d}{d\alpha} f(A+\alpha U)\Big|_{\alpha=0} = \frac{d}{d\alpha} \cdot u^T(A+\alpha U)v\Big|_{\alpha=0}$$

$$= u^T U v = \langle \underset{m\times n}{\nabla}, \underset{m\times n}{U} \rangle$$

$$u^T U v = u^T(Uv) = \langle u, Uv \rangle = \langle uv^T, U \rangle$$

$$\nabla = \underbrace{u}_{m\times 1}\underbrace{v^T}_{1\times n} \in \mathbb{R}^{m\times n}$$

Linear Regression

K. N. Toosi
University of Technology

$$\overset{\in\mathbb{R}^n}{x} \longrightarrow \boxed{\;f\;} \longrightarrow y \in \mathbb{R}^m \qquad y = f(x) = \underset{\in\mathbb{R}^{m\times n}}{A}\,\overset{\in\mathbb{R}^n}{x} + \underset{\in\mathbb{R}^m}{b} \in \mathbb{R}^m$$

$$(\underset{\mathbb{R}^n}{\vec{x}_1}, \underset{\mathbb{R}^m}{\vec{y}_1}), (x_2, y_2), \dots, (x_p, y_p) \qquad f(x; A, b)$$

$$\cancel{Ax_i \simeq y_i} \quad Ax_i + b \simeq y_i$$

$$Ax_i + \vec{b} - \vec{y}_i \in \mathbb{R}^m$$

$$\|Ax_1 + b - y_1\|^2 + \|Ax_2 + b - y_2\|^2 + \dots + \|Ax_p + b - y_p\|^2$$

$$C(\theta) = C(A, b) = \sum_{i=1}^{p} \|Ax_i + b - y_i\|^2$$ cost function

$$\theta = (A, b)$$

$$A^*, b^* = \arg\min C(A, b)$$

$$C: \mathbb{R}^{m\times n} \times \mathbb{R}^m \longrightarrow \mathbb{R}$$

Linear Regression

$$\frac{\partial C^T}{\partial b} \longrightarrow \text{gradient w.r.to } b$$

$$\nabla^{b}_{f_b} = \frac{\partial C}{\partial b} = \sum_{i=1}^{p} \frac{\partial}{\partial b} \|Ax_i + b - y_i\|^2 = \sum_{i=1}^{p} \frac{\partial}{\partial b} (Ax_i - y_i + b)^T \underline{(Ax_i - y_i + b)}$$

$$= \sum_{i=1}^{p} 2(Ax_i - y_i + b) = 0$$

$$\Rightarrow A\sum x_i - \sum y_i + pb = 0 \Rightarrow b^* = -A \underbrace{\boxed{\frac{1}{p}\sum_{i=1}^{p} x_i}}_{\mu_x} + \underbrace{\frac{1}{p}\sum y_i}_{\mu_y}$$

$$\boxed{b^* = -A\mu_x + \mu_y}$$

$$C(A,b) = \sum_{i=1}^{p} \|Ax_i - y_i + b\|^2 = \sum_{i=1}^{p} \|Ax_i - y_i - A\mu_x + \mu_y\|^2$$

$$= \sum_{i=1}^{p} \|A\underbrace{(x_i - \mu_x)}_{\bar{x}_i} - \underbrace{(y_i - \mu_y)}_{\bar{y}_i}\|^2 = \sum_{i=1}^{p} \|\underline{A}\,\bar{x}_i - \bar{y}_i\|^2 = C'(A)$$

$$\underline{C' : \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}} \qquad \frac{\partial C'(A)}{\partial A} = \nabla^{A}_{C'} = 0$$

Linear Regression

III

$$C'(A) = \sum_{i=1}^{p} \|A\bar{x}_i - \bar{y}_i\|^2$$

$$= \sum_{i=1}^{p} (A\bar{x}_i - \bar{y}_i)^T (Ax_i - y_i)$$

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

$$= 2\sum_{i=1}^{p} (Ux_i)^T (Ax_i - y_i)$$

$$= 2\sum_{i=1}^{p} \langle Ux_i , Ax_i - y_i \rangle$$

$$= 2\sum_{i=1}^{p} \langle U , \underbrace{\underbrace{(Ax_i - y_i)}_{m\times 1} \underbrace{x_i^T}_{1\times n}}_{m\times n} \rangle$$

$\langle\cdot,\cdot\rangle$ bilinear

$$= \langle U , \underbrace{2\sum_{i=1}^{p} (Ax_i - y_i) x_i^T}_{\nabla_{C'}^{A} = \frac{\partial C'}{\partial A} \in \mathbb{R}^{m\times n}} \rangle$$

$$\underbrace{XX^T} \qquad \underbrace{YX^T}$$

Linear Regression

$$= \Big\langle U \,,\, \underbrace{2\sum_{i=1}^{p} (A x_i - y_i)\, x_i^T}_{\nabla_{C'}^{A} = \frac{\partial C'}{\partial A} \in \mathbb{R}^{m\times n}} \Big\rangle$$

$$\nabla = 2\sum_{i=1}^{p} \underbrace{A}_{m\times n}\,\underbrace{x_i}_{n\times 1}\,\underbrace{x_i^T}_{1\times n} - \underbrace{y_i}_{m\times 1}\,\underbrace{x_i^T}_{n\times n} = 2A\overbrace{\sum_{i=1}^{p} x_i x_i^T}^{XX^T} - 2\overbrace{\sum_{i=1}^{p} y_i x_i^T}^{YX^T}$$

$$X = [x_1, x_2, \dots, x_p] \in \mathbb{R}^{n\times p}, \quad Y = [y_1, y_2, \dots, y_p] \in \mathbb{R}^{m\times p}$$

$$\nabla = 2A\,XX^T - 2\,YX^T \qquad \nabla = 0 \Rightarrow A_{opt} = YX^T(XX^T)^{-1}$$

$$= \Big(\sum_{i=1}^{p} y_i x_i^T\Big)\Big(\sum_{i=1}^{p} x_i x_i^T\Big)^{-1}$$

XX^T invertible (non-singular) $\Longleftrightarrow$ X full row rank $\Rightarrow$ rank$(X) = m$

$\Rightarrow \exists$ m independent columns $\Rightarrow$ span$(x_1, x_2, \dots, x_n) \in \mathbb{R}^m$

$A = \checkmark$

Linearization

K. N. Toosi
University of Technology

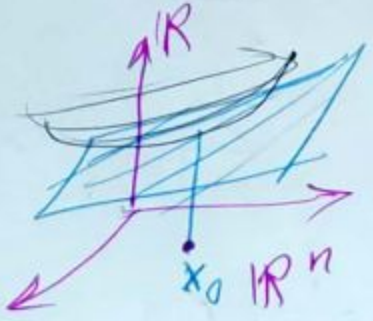
MA 17

$f(x) \longrightarrow$ linearize around x_0

$$f(x) \simeq f(x_0) + \nabla_f(x_0)^T (x - x_0)$$

$$\nabla^T x + b$$

$b \rightarrow f(x_0) - \nabla^T x_0$



Vector-valued multivariate functions

K. N. Toosi
University of Technology

$\mathbb{R} \to \mathbb{R} \quad \to$

$\mathbb{R} \to \mathbb{R}^n \quad \mathbb{R} \to \mathbb{R}^{n\times m}$ vector valued, matrix valued of one variable

$\mathbb{R}^n \to \mathbb{R} \quad \mathbb{R}^{n\times m} \to \mathbb{R}$ multivariate function

---

$f: \mathbb{R}^n \to \mathbb{R}^m$ vector valued, multi variate

Vector-valued multivariate functions: linearization

$f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ vector valued, multi variate

$$f(x) = \underbrace{f(x_0)}_{\mathbb{R}^m} + \underbrace{M}_{m\times n}\underbrace{(x - x_0)}_{\mathbb{R}^n}$$

$m = 1 \Rightarrow M = \nabla^T$

transposed gradient

Jacobian

The Jacobian matrix

K. N. Toosi
University of Technology

$f: \mathbb{R}^n \to \mathbb{R}^m$ vector valued, multi variate

$$f(X) = \underbrace{f(X_0)}_{\mathbb{R}^m} + \underbrace{M}_{m\times n}\underbrace{(X - X_0)}_{\mathbb{R}^n} \qquad m=1 \Rightarrow M = \nabla^T$$

transposed gradient

$\boxed{M = J}$ یعقوبی ژاکوبین Jacobian

$\cancel{f(X)}$ $f: \mathbb{R}^n \to \mathbb{R}^m$

$$Y = f(X) = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_m(X) \end{bmatrix}$$

$$f(X) = \begin{bmatrix} f_1(X) \\ f_2(X) \\ \vdots \\ f_m(X) \end{bmatrix} \qquad f_i: \mathbb{R}^n \to \mathbb{R}$$

$$\text{Jacobian } f = J_f = \begin{bmatrix} \nabla f_1^T \\ \nabla f_2^T \\ \vdots \\ \nabla f_m^T \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

The Jacobian matrix: Example

K. N. Toosi
University of Technology

$$f(x) = A\mathbf{x} \qquad A \in \mathbb{R}^{m\times n}$$

$$f_{i:} : a_{i:}^T x = \sum_{k=1}^{n} a_{ik} x_k$$

$$(J_f)_{ij} = \frac{\partial f_i}{x_j} = a_{ij} \Rightarrow J_f = A$$

Calculating the Jacobian: method 1

Calculate Jacobian: Method 1 MA17

$$(J_f)_{ij} = \frac{\partial f_i}{\partial x_j} \longrightarrow \text{arrange in a matrix \& simplify}$$

Calculating the Jacobian: method 2

K. N. Toosi
University of Technology

Method 2: directional derivative

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

$$D[u]f(x) = \frac{d}{d\alpha} f(x + \alpha u)\Big|_{\alpha=0} \in \mathbb{R}^m$$

linear w.r.t u

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

Calculating the Jacobian: method 2 - Example

K. N. Toosi
University of Technology

$$f(x) = Ax$$

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

$$= Au$$

J