

MATRIX CALCULUS USING "THE TRACE TRICK"

Trace: sum of diagonal Elements

Properties :

- $\text{Tr}[A+B] = \text{Tr}[A] + \text{Tr}[B]$
- $\text{Tr}[cA] = c \text{Tr}[A]$
- $\text{Tr}[AB] = \text{Tr}[BA]$
- $\text{Tr}[A] = \text{Tr}[A^T]$

$$\text{Tr}(A) = a_{11} + a_{22} + \dots + a_{nn}$$

Differential: small change in all elements.

Properties :

- $d(cA) = c dA$
- $d(A+B) = dA + dB$
- $d(AB) = dA B + B dA$

$$dA = \begin{bmatrix} da_{11} & \dots & da_{1n} \\ \vdots & & \vdots \\ da_{m1} & \dots & da_{mn} \end{bmatrix}$$

Theorem :

If we can write differential of scalar function $L(x)$ into the form:

$$dL = \text{Tr}(G^T dX)$$

Then gradient of L w.r.t X is :

$$\nabla_x L = 6$$

Problem 4 : Trace of linear product

Let $L = \text{Tr}(AX)$. Find $\nabla_x L$

Problem 4 : Trace of linear product

Let $L = \text{Tr}(AX)$. Find $\nabla_x L$

Step 1: Take Differential on both sides

$$dL = d(\text{Tr}(AX))$$

Step 2. Since trace is sum and d is a linear operator, we can swap d and Tr .

$$dL = \text{Tr}(d(AX)) = \text{Tr}(A dX)$$

Step 3. Identify the matrix G .

$$G^T = A$$

$$\Rightarrow G = A^T$$

Answer: $\nabla_x L = A^T$

Problem 5 : Trace of Quadratic product

Let $L = \text{tr}(x^T A x)$, where A is constant sq. matrix. Find $\nabla_x L$.

Problem 5: Trace of Quadratic product

Let $L = \text{Tr}(x^T A x)$, where A is constant sq. matrix. Find $\nabla_x L$.

Step 1: Take differential on both sides & use product rule of differential.

$$\begin{aligned} dL &= \text{Tr}(d(x^T A x)) \\ &= \text{Tr}(dx^T A x + x^T A dx) \end{aligned}$$

Step 2: Linearity of Trace

$$dL = \text{Tr}(dx^T A x) + \text{Tr}(x^T A dx) \quad \text{--- (1)}$$

Step 3: Trace of transposed is same.

$$\text{Tr}(dx^T A x) = \text{Tr}(x^T A^T dx) \quad \text{--- (11)}$$

using it in (1):

$$\begin{aligned} dL &= \text{Tr}(x^T A^T dx) + \text{Tr}(x^T A dx) \\ &= \text{Tr}(x^T A^T dx + x^T A dx) \\ &= \text{Tr}(x^T (A^T + A) dx) \end{aligned}$$

Step 4: Identify G .

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

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

Answer: $\nabla_x L = (A + A^T)x$

Problem 6: Linear Regression

Let $L = \|Ax - b\|_2^2$. Derive $\nabla_x L$ and set it to 0.

Step 1. Convert L to Trace form.

$$L = (Ax - b)^T (Ax - b)$$

$$= R^T R$$

$$= \text{Tr}(R^T R) \quad \text{where } R = Ax - b$$

Step 2. Apply differential on both sides

$$dL = \text{Tr}(dR^T R + R^T dR)$$

$$= \text{Tr}(2R^T dR) \quad \left(\begin{array}{l} \text{remember} \\ \text{problem 5} \end{array} \right)$$

Step 3. Find dR .

$$\begin{aligned} dR &= d(AX - b) \\ &= A dX - \underbrace{db}_{\rightarrow 0} \\ &= A dX \end{aligned}$$

Plug it back to dL :

$$\begin{aligned} dL &= \text{Tr}(2 R^T A dX) \\ &= \text{Tr}(2 (AX - b)^T A dX) \end{aligned}$$

Step 4: Identify G .

$$G^T = 2 (AX - b)^T A$$

$$G = 2 A^T (AX - b)$$

$$\nabla_X L = 2 A^T (AX - b)$$

Step 5: Set it to zero

$$X = (A^T A)^{-1} A^T b$$