MATRIX CALCULUS USING "THE TRACE TRICK"

Trace : sum of diagonal Elements

Properties :

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

$$Tr(A) = \begin{bmatrix} a_{11} & & & \\ & a_{12} & & \\ & & \ddots & \\ & & & a_{mn} \end{bmatrix}$$

Differential : small change in all elements.

Properties :

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

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

Theorem :

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

$$dL = Tr(G^T dX)$$

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

$$\nabla_X L = G$$

Problem 4 : Trace of linear product

Let $L = \mathrm{Tr}(AX)$. Find $\nabla_X L$

Problem 4 : Trace of linear product

Let $L = Tr(AX)$. Find $\nabla_X L$

Step 1: Take Differential on both sides

$$dL = d(Tr(AX))$$

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

$$dL = Tr(d(AX)) = Tr(A\,dX)$$

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 $L = tr(X^T A X)$, where A is constant sq. matrix. Find $\nabla_X L$.

Problem 5: Trace of Quadratic product

Let $L = Tr(X^T A X)$, where A is constant sq. matrix. Find $\nabla_X L$.

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

$$dL = Tr(d(X^T A X))$$

$$= Tr(dX^T A X + X^T A dX)$$

Step 2: Linearity of Trace

$$dL = Tr(dX^T A X) + (X^T A dX) \quad (1)$$

Step 3: Trace of transposed is same.

$$Tr(dX^T A X) = Tr(X^T A^T dX) \quad (11)$$

using it in (1):

$$dL = Tr(X^T A^T dX) + Tr(X^T A dX)$$

$$= Tr(X^T A^T dX + X^T A dX)$$

$$= Tr(X^T (A^T + A) dX)$$

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 $= 0$.

Step 1. Convert L to Trace form.

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

$$= R^T R$$

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

Step 2. Apply differential on both sides

$$dL = Tr(dR^T R + R^T dR)$$

$$= Tr(2R^T dR)$$ (remember problem 5)

Step 3. Find dR.

$$dR = d(AX - b)$$
$$= A\,dX - \underbrace{db}_{\to\, 0}$$
$$= A\,dX$$

Plug it back to dL:

$$dL = Tr(2R^T A\,dX)$$
$$= Tr(2(AX-b)^T A\,dX)$$

Step 4: Identify G.

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

Step 5: Set it to zero

$$\boxed{X = \frac{1}{2}(A^TA)^{-1}A^Tb}$$