cs229 notes

Bhavit Sharma

4/24/23

Table of contents

1 Matrix Derivatives

Good link: https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

1.1 Some useful identities

1.

$$\frac{\partial}{\partial X} \log |X| = X^{-1}$$

Proof:

$$\frac{\partial}{\partial X} \log |X| = \frac{1}{|X|} \frac{\partial |X|}{\partial X}$$

We know that

$$(\frac{\partial |X|}{\partial X})_{ij} = \frac{\partial}{\partial X_{ij}} * det(X)$$

and

$$det(X) = X_{i1}C_{i1} + X_{i2}C_{i2} + \cdots + X_{in}C_{in}$$

where C_{ij} is the cofactor of X_{ij}. So,

$$\frac{\partial}{\partial X_{ij}} * det(X) = C_{ij}$$

$$\frac{\partial |X|}{\partial X} = C = adj(X)^T$$

where C is the cofactor matrix of X. $adj(X)$ is the adjugate matrix of X and $X^{-1} = \frac{adjX}{|X|}$.

so we get

$$\frac{\partial}{\partial X} \log |X| = \frac{1}{|X|} \frac{\partial |X|}{\partial X} = \frac{1}{|X|} adj(X)^T = (X^{-1})^T$$

Reference: kamper matrix calculus

2. $\frac{\partial}{\partial X}(z^T X^{-1} z) = -(X^{-1}) z z^T (X^{-1})$

Proof:

$$\frac{\partial}{\partial X}(z^T X^{-1} z)$$

Lets first compute the derivative of $z^T X^{-1} z$ with respect to X_{ij}

$$\frac{\partial}{\partial X_{ij}}(z^T X^{-1} z)$$

Lets first derive $\frac{\partial X^{-1}}{\partial X_{ij}}$

$$\frac{\partial X^{-1}}{\partial X_{ij}}$$

Using $X * X^{-1} = I$ we get

$$X^{-1} \frac{\partial X}{\partial X_{ij}} + \frac{\partial X^{-1}}{\partial X_{ij}} X = 0$$

i.e.

$$\frac{\partial X^{-1}}{\partial X_{ij}} = -X^{-1} \frac{\partial X}{\partial X_{ij}} X^{-1}$$

where $\frac{\partial X}{\partial X_{ij}}$ is the matrix of partial derivatives of X with respect to X_{ij} and it's elements are 0 except for the element at i, j which is 1.

So lets say $H = \frac{\partial \ tr(z^T X^{-1} z)}{\partial X}$

$$H_{ij} = \frac{\partial}{\partial X_{ij}} tr(z^T X^{-1} z)$$

Using cyclic property of trace we get

$$H_{ij} = \frac{\partial}{\partial X_{ij}} tr(z^T X^{-1} z) = \frac{\partial}{\partial X_{ij}} tr(zz^T (X^{-1}))$$

We know that

$$\partial(Tr(A)) = Tr(\partial(A))$$

because trace is linear. so

$$H_{ij} = tr(zz^T \frac{\partial}{\partial X_{ij}}(X^{-1})) = tr(zz^T(-X^{-1}\frac{\partial X}{\partial X_{ij}} X^{-1}))$$

Using cyclic property of trace we get

$$H_{ij} = tr(X^{-1} zz^T X^{-1} \frac{\partial X}{\partial X_{ij}})$$

Now suppose that

$$F = X^{-1} zz^T X^{-1}$$

then

$$tr(F\frac{\partial X}{\partial X_{ij}}) = F_{ji} = F_{ij}$$

since F is symmetric. Hint: You can think of the fact only the jth row of F is multiplied by the jth column, and only ith column of jth row of F is multiplied by the ith row of jth column of F leading to element at F_{jj} contributing and the rest being zero.

Hence: $H = -X^{-1} zz^T X^{-1}$

2 Prove that if $z^T H z \geq 0$ then H is positive semi-definite and cost function J is convex. H is Hessian matrix of J.