Matrix Calculus

Date
Page

1. Gradient

⇒ Suppose that $f: \mathbb{R}^{m \times n} \to \mathbb{R}$ is a function that takes as input a matrix A of size $m \times n$ and returns a real value.

⇒ Then the gradient of f (with respect to A) is the matrix of partial derivatives, defined as:

$$\nabla_A f(A) \in \mathbb{R}^{m \times n} = \begin{bmatrix} \frac{\delta f(A)}{\delta A_{11}} & \frac{\delta f(A)}{\delta A_{12}} & \cdots & \frac{\delta f(A)}{\delta A_{1n}} \\ \vdots & \vdots & & \vdots \\ \frac{\delta f(A)}{\delta A_{m1}} & \frac{\delta f(A)}{\delta A_{m2}} & \cdots & \frac{\delta f(A)}{\delta A_{mn}} \end{bmatrix}$$

or

$$\left(\nabla_A f(A)\right)_{ij} = \frac{\delta f(A)}{\delta A_{ij}}$$

⇒ If, in particular, A is just a vector $x \in \mathbb{R}^n$

$$\nabla_x f(x) = \begin{bmatrix} \frac{\delta f(x)}{\delta x_1} \\ \frac{\delta f(x)}{\delta x_2} \\ \vdots \\ \frac{\delta f(x)}{\delta x_n} \end{bmatrix}$$

Date
Page

⇒ It is very important to remember that the gradient of a function is only defined if the function is real-values, that is, if it returns a scalar value.

$$\nabla_x (af(x) + bg(x)) = a\nabla_x f(x) + b\nabla_x g(x)$$

2. The Hessian

⇒ Suppose that $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a function that takes a vector in $\mathbb{R}^n$ and returns a real number.

⇒ Then the Hessian matrix with respect to x, written $\nabla_x^2 f(x)$ or simply H is the $n \times n$ matrix of partial derivatives

$$\nabla_x^2 f(x) \in \mathbb{R}^{n \times n} = \begin{bmatrix} \frac{\delta^2 f(x)}{\delta x_1^2} & \frac{\delta^2 f(x)}{\delta x_1 \delta x_2} & \cdots & \frac{\delta^2 f(x)}{\delta x_1 \delta x_n} \\ \frac{\delta^2 f(x)}{\delta x_2 \delta x_1} & \frac{\delta^2 f(x)}{\delta x_2^2} & \cdots & \frac{\delta^2 f(x)}{\delta x_2 \delta x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\delta^2 f(x)}{\delta x_n \delta x_1} & \frac{\delta^2 f(x)}{\delta x_n \delta x_2} & \cdots & \frac{\delta^2 f(x)}{\delta x_n^2} \end{bmatrix}$$

or

$$\left(\nabla_x^2 f(x)\right)_{ij} = \frac{\delta^2 f(x)}{\delta x_i \delta x_j}$$

⇒ Hessian is always symmetric, since

$$\frac{\delta^2 f(x)}{\delta x_i \delta x_j} = \frac{\delta^2 f(x)}{\delta x_j \delta x_i}$$

⇒ Similar to the gradient, the Hessian is defined only when $f(x)$ is real-valued.

⇒ Gradient of gradient is not hessian for a vector valued function.

↳ The gradient of the function is a vector, and we cannot take the gradient of a vector.

⇒ If we don't mind being a little bit sloppy we can say that

$$\nabla_x^2 f(x) = \nabla_x \left(\nabla_x f(x) \right)^T$$

→ As long as we understand that this really means taking the gradient of each entry of $(\nabla_x f(x))^T$, not the gradient of the whole vector.

Date
Page

3. Gradient and Hessians of Quadratic & Linear functions

1) $\nabla_x b^T x = b$

2) $\nabla_x x^T A x = (A + A^T)x \longrightarrow 2Ax$ {if A symmetric}

3) $\nabla_x^2 x^T A x = (A + A^T) \longrightarrow 2A$ {if A symmetric}

4) $\nabla_A |A| = (adj(A))^T = |A| A^{-T}$

5) $\nabla_A \log|A| = \frac{1}{|A|} \nabla_A |A| = A^{-T}$