Foundation of Machine Learning

Revisiting Prerequisite of Maths : Calculus

Sayanta Adhikari

MTech(RA)
Department of Artificial Intelligence
Indian Institute of Technology, Hyderabad

Table of Contents

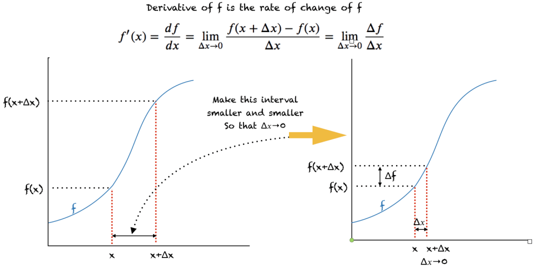
Table of Contents

Derivatives

The slope of the secent line through two points on the graph of f is given by :

$$\frac{\delta f(x)}{\delta x} = \frac{f(x + \delta x) - f(x)}{\delta x}$$

As we reduce δx we reach a limiting case where we get the slope at a particular point.



The derivative of f points in the direction of steepest ascent of f.

Partial Derivatives

Lets consider $f : \mathbb{R}^n \to \mathbb{R}$. So finding partial derivative comes into picture when you would like to find the derivative of the function w.r.t. each element/dimension of input $\mathbf{x}$.

$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1 + h, x_2, \cdots, x_n) - f(\mathbf{x})}{h}$$

$$\vdots$$

$$\frac{\partial f}{\partial x_n} = \lim_{h \to 0} \frac{f(x_1, x_2, \cdots, x_n + h) - f(\mathbf{x})}{h}$$

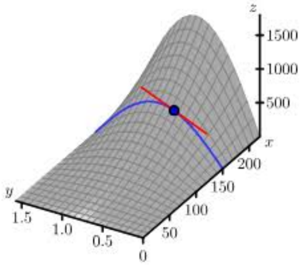


Table of Contents

Gradients

Gradients are derivatives of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ w.r.t input vector x. It is exactly equal to the collection of all the partial derivatives in a particular order.

$$\nabla_x f = grad\ f = \frac{df}{dx} = [\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \cdots, \frac{\partial f(x)}{\partial x_n}] \in \mathbb{R}^{1\times n}$$

For this session we consider gradients to be Row Vectors. But its not uncommon to represent gradients as column vectors.

Chain Rule:

$$\frac{\partial}{\partial x}(g \circ f)(x) = \frac{\partial g}{\partial f}\frac{\partial f}{\partial x}$$

In this case also its the same as it was for uni-variate case.

Jacobian

Let's make derivatives more general. Consider a function $f : \mathbb{R}^n \to \mathbb{R}^m$ with input as $x = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$ and output of it is given by a vector function

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_m(x) \end{bmatrix} \in \mathbb{R}^m$$

Then the derivative of f w.r.t. x is given by the collection of all possible partial derivatives arranged as a matrix $J \in \mathbb{R}^{m \times n}$ in a certain order.

$$J = \frac{df(x)}{dx} = \begin{bmatrix} \nabla_x f_1 \\ \nabla_x f_2 \\ \vdots \\ \nabla_x f_m \end{bmatrix} \textit{ where } \nabla_x f_i \in \mathbb{R}^n$$

Example

Consider a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ with input as $x(t) = [x_1(t), x_2(t)]^T$ which is itself a function $x : \mathbb{R} \rightarrow \mathbb{R}^2$. We need to find $\frac{df}{dt}$.

Example

Going beyond 2-Dimension Jacobian

Consider a function $f : \mathbb{R}^{m\times n} \to \mathbb{R}^{p}$. A function that takes matrix as input and outputs a vector.
Let the input be $A \in \mathbb{R}^{m\times n}$ and the output be $f(A) = [f_1(A), \cdots, f_p(A)]^T$
Then to find the derivative, its again the collection of all possible partial derivatives presented in a particular order.

$$\frac{df}{dA} = \begin{bmatrix} \frac{df_1(A)}{dA} \\ \frac{df_2(A)}{dA} \\ \vdots \\ \frac{df_p}{dA} \end{bmatrix}, \frac{df_i}{dA} \in \mathbb{R}^{1\times(m\times n)}$$

Similarly, one can find dimension of the derivatives as well as the particular ordering even for higher dimensions like $\mathbb{R}^{m\times n} \to \mathbb{R}^{p\times q}$,
$\mathbb{R}^{m\times n\times k} \to \mathbb{R}^{p\times q}$and so on.

Table of Contents

Some of the most seen functions in ML

- $\frac{\partial x^T a}{\partial x} = a^T$
- $\frac{\partial a^T x}{\partial x} = a^T$
- $\frac{\partial a^T X b}{\partial X} = ab^T$
- $\frac{\partial x^T B x}{\partial x} = x^T(B + B^T)$
- $\frac{\partial (x-As)^T W (x-As)}{\partial s} = -2(x - As)^T WA$ for symmetric W
- $\frac{\partial a^T X^{-1} b}{\partial X} = -(X^{-1})^T ab^T (X^{-1})^T$