# Machine Learning for Engineering and Science Applications

Derivatives

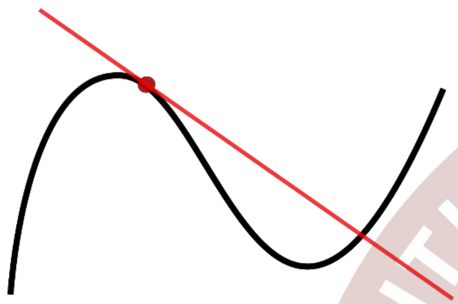
Gradient

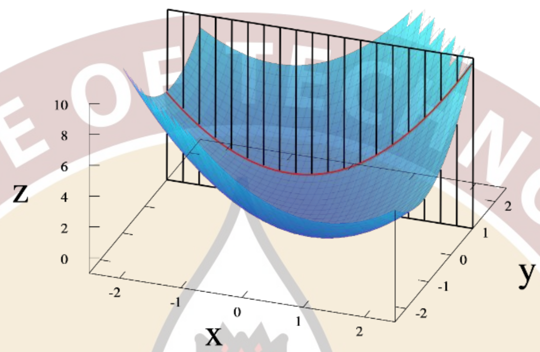
Hessian

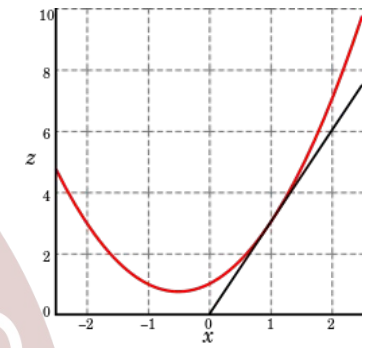
Jacobian

Taylor Series

# Derivatives






- Derivatives measure how one much quantity changes when there is a small change in another

- Geometrically, in one dimension, this can be given as the slope of the tangent

$$f'(a) = \frac{df}{dx}(x = a) = \lim_{h\to 0} \frac{f(x+h) - f(x)}{h}$$

- In higher dimensions (functions of many variables/vectors), we have the idea of partial derivatives

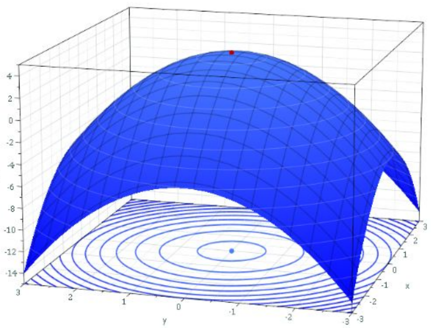
$$\frac{\partial f}{\partial x_i}(a_1, \ldots, a_n) = \lim_{h\to 0} \frac{f(a_1, \ldots, a_i + h, \ldots, a_n) - f(a_1, \ldots, a_i, \ldots, a_n)}{h}$$

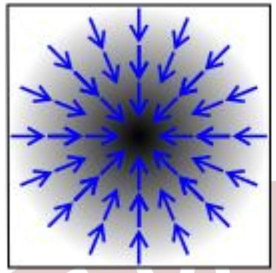
https://en.wikipedia.org/wiki/Derivative#/media/File:Tangent_to_a_curve.svg

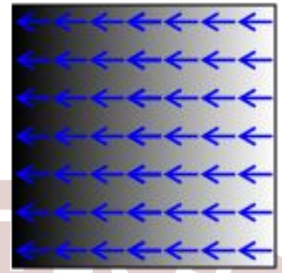
https://en.wikipedia.org/wiki/Partial_derivative

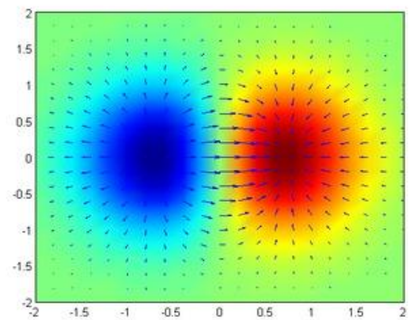
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

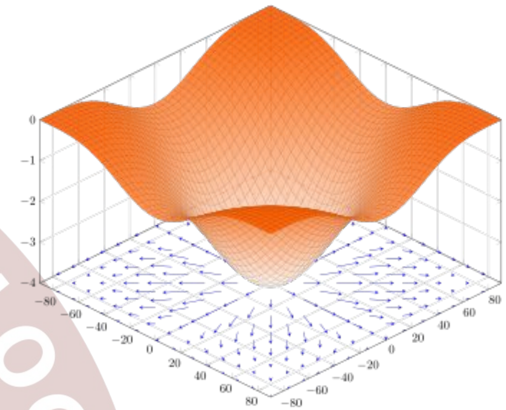
# Gradient











- The gradient is the multivariable generalization of the derivative

- It is a vector the components of which denote the partial derivatives in each direction

$$\nabla_{\mathbf{x}} f(x_1, \dots, x_n) = \left[\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}\right]^T$$

- It can be used to calculate the directional partial derivative of $f$ along the direction $\hat{\boldsymbol{v}}$

$$D_v f(\mathbf{x}) = \lim_{\alpha \to 0} \frac{\partial f(\mathbf{x} + \alpha \hat{\boldsymbol{v}})}{\partial \alpha} = \nabla_{\mathbf{x}} f(\mathbf{x}) \cdot \hat{\boldsymbol{v}}$$

https://en.wikipedia.org/wiki/Maxima_and_minima#/media/File:MaximumParaboloid.png

https://en.wikipedia.org/wiki/Gradient

# Hessian

- The Hessian is the gradient of the gradient.
  - It is the equivalent of the second derivative in scalar calculus and has similar uses

- For $f: \mathbb{R}^n \to \mathbb{R}$, we have $H_{i,j} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ is the Hessian which is a $n \times n$ matrix

$$H_{i,j} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- Note that the Hessian is a symmetric matrix

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

# Jacobian

- The Jacobian is the equivalent of the gradient for vector valued functions
  - The Hessian can be seen as the gradient (Jacobian) of a gradient (which is a vector)

- For $\boldsymbol{f}: \mathbb{R}^n \rightarrow \mathbb{R}^m$, we have $J_{i,j} = \frac{\partial^2 f(\boldsymbol{x})_i}{\partial x_j}$ is the Jacobian which is a J $\in \mathbb{R}^{m \times n}$

$$J = \nabla_{\mathbf{x}} \boldsymbol{f} = \begin{bmatrix} \frac{\partial^2 f_1}{\partial x_1} & \frac{\partial^2 f_1}{\partial x_2} & \cdots & \frac{\partial^2 f_1}{\partial x_n} \\ \frac{\partial^2 f_2}{\partial x_1} & \frac{\partial^2 f_2}{\partial x_2} & \cdots & \frac{\partial^2 f_2}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f_m}{\partial x_1} & \frac{\partial^2 f_m}{\partial x_2} & \cdots & \frac{\partial^2 f_m}{\partial x_n} \end{bmatrix} \in \mathbb{R}^{m \times n}$$

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

# Taylor Series

- The Taylor series is a local approximation of a function's value in terms of polynomials
  - It is an extremely useful and widely used idea in multiple fields
  - There are mathematical subtleties which we will be ignoring here

- For $f: \mathbb{R} \rightarrow \mathbb{R}$, recall that the Taylor series is written as

$$f(x) \approx f(x^0) + (x - x^0)\frac{df}{dx} + \frac{1}{2}(x - x^0)^2\frac{d^2f}{dx^2} + \cdots$$

- For $f: \mathbb{R}^n \rightarrow \mathbb{R}$, the Taylor series can be written as

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}^0) + (\boldsymbol{x} - \boldsymbol{x}^0)^T\boldsymbol{g} + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^0)^T\boldsymbol{H}(\boldsymbol{x} - \boldsymbol{x}^0) + \cdots$$

- Here, $\boldsymbol{g} = \nabla_x f(\boldsymbol{x}^0)$ and $\boldsymbol{H}$ is the Hessian calculated at $\boldsymbol{x}^0$ also