



**BUILD YOUR  
SOLID MATH  
FOUNDATION**



# HELLO!

## I am Cao Thanh Hà

I am here because I love to give presentations.

You can find me at [caothanhha9@yahoo.com/gmail.com](mailto:caothanhha9@yahoo.com/gmail.com)

A thick yellow diagonal stripe runs from the top right corner towards the bottom left, separating the white background on the left from the solid yellow background on the right.

1.

WHY MATH?

Tại sao chúng ta  
cần có bài học  
này?

“

Các mô hình machine learning đều được biểu diễn bằng các công thức toán học.



## TODAY'S TASKS

- ▶ Đại số tuyến tính: không gian vector và tensor
- ▶ Giải tích: đạo hàm và gradient descent
- ▶ Xác suất thống kê: công thức bayes và thống kê

Trong bài học này bạn sẽ thấy những công thức quen thuộc được sử dụng trong machine learning



# LINEAR ALGIBRA

Không gian vector

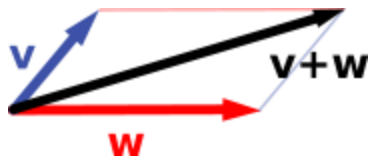
Scalar  $\rightarrow$  vector  $\rightarrow$  matrix  $\rightarrow$  tensor



## Không gian **VECTOR**

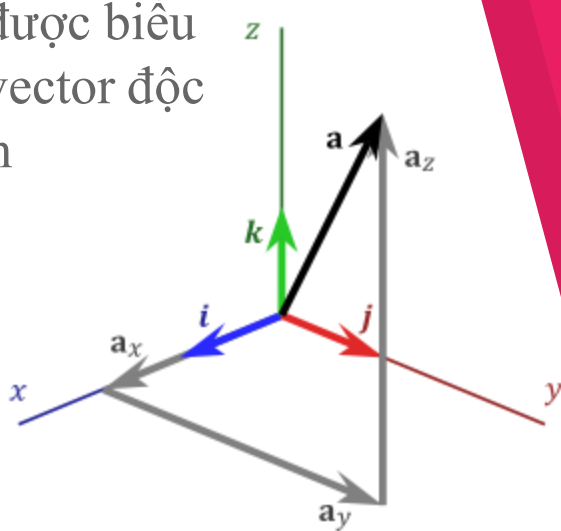
### 2D space

Không gian được biểu diễn bằng 2 vector độc lập tuyến tính



### 3D space

Không gian được biểu diễn bằng 3 vector độc lập tuyến tính

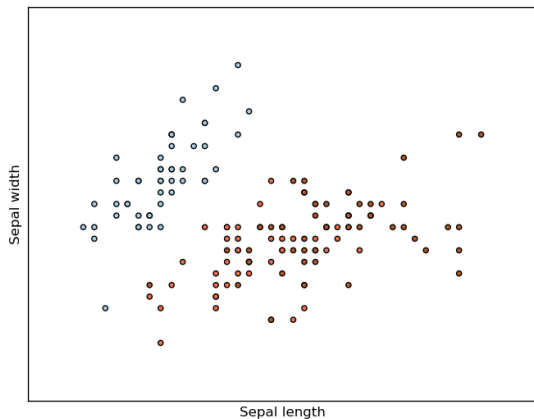




# Không gian **VECTOR** và machine learning

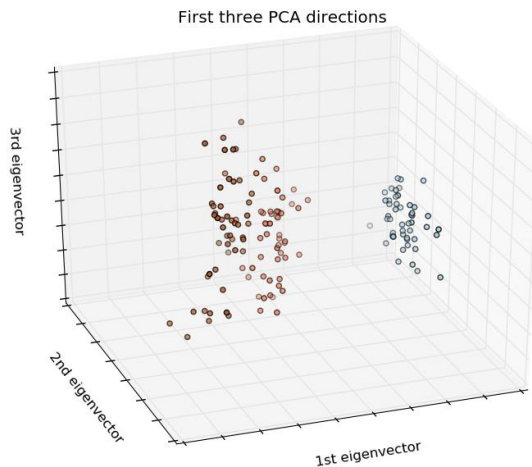
## 2D space

Dữ liệu có thể có 2 chiều



## 3D space

Dữ liệu có thể là 3 chiều







# Scalar, vector, matrix và tensor

$$x=1$$

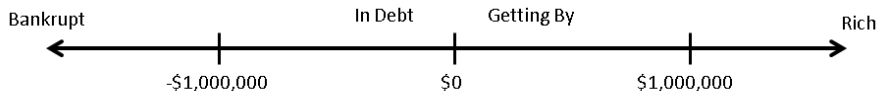
$$\mathbf{x}=[0,1]$$

$$\mathbf{X}=[[1,2],[3,4],[4,5]]$$

$$\mathbf{X}=[[1,2],[3,4],[4,5]], [[6,7],[8,9],[10,11]]$$

The Income Line

M. Strassler 2011



## Training

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1D} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2D} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3D} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{N1} & x_{N2} & x_{N3} & \cdots & x_{ND} \end{bmatrix}$$

Random Feature Selection  $\mathcal{R}^m$

$$\mathbf{X}^m = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1B} \\ x_{21} & x_{22} & \cdots & x_{2B} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NB} \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1C} \\ y_{21} & y_{22} & \cdots & y_{2C} \\ \vdots & \vdots & \cdots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NC} \end{bmatrix}$$

CCA  $\rightarrow (\mathbf{W}^m, \mathbf{o}^m)$

## Testing

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1D} \end{bmatrix}$$

Selected Features  $\mathcal{R}^m$

$$\mathbf{X}^m = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1B} \end{bmatrix}$$

$$\mathbf{b}^m = (\mathbf{X}^m \mathbf{W}^m + \mathbf{o}^m) > 0$$

$\mathbf{b}^m$

$p_{l,c_i} = \frac{N_{l,c_i}}{N_{c_i}}$



# MATHEMATICAL ANALYSIS

Hàm số

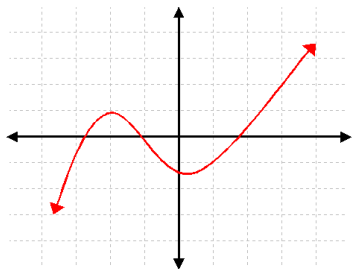
Phép tính đạo hàm và gradient descent



# Hàm số - PHÉP ÁNH XẠ

## Hàm một biến

Ánh xạ một số từ tập này sang tập khác



## Hàm nhiều biến

Ánh xạ từ không gian vector sang không gian vector

Function Notation

$$y = f(x)$$

Diagram illustrating the function notation  $y = f(x)$  with labels:

- Output**: Points to  $y$
- Name of Function**: Points to  $f$
- Input**: Points to  $x$

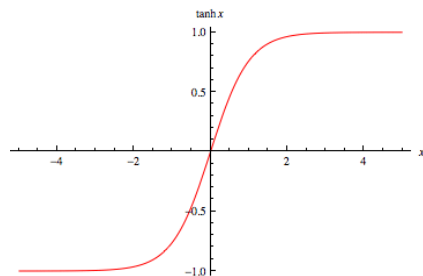
## Toán tử

Ánh xạ một tensor từ một không gian này sang một không gian khác

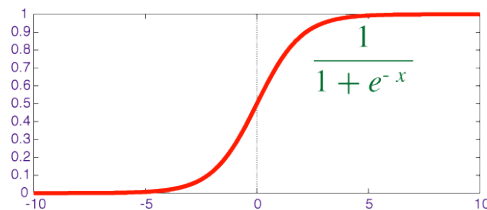
Trong machine learning, hàm số là phép biến đổi input (data) thành output (classification, regression)

# Các hàm số gấ trong ML (deep learning)

► Hàm tanh



► Hàm sigmoid



► Hàm soft-max

$$p(i) = \frac{e^{\frac{f(i)}{T}}}{\sum_j e^{\frac{f(j)}{T}}}$$

$$\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



# Đạo hàm – GRADIENT DESCENT

## Đạo hàm một biến

Giới hạn của biến thiên output theo biến thiên input

$$\frac{dy}{dx} = \lim_{\delta x \rightarrow 0} \frac{f(x + \delta x) - f(x)}{\delta x}$$

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial u} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial u}$$

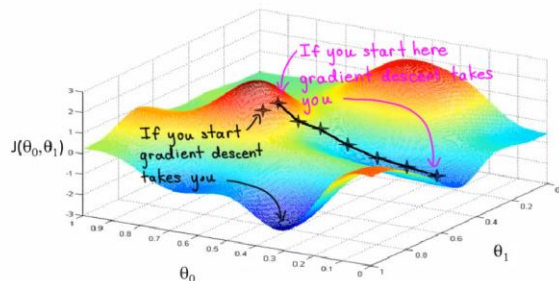
$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial v}$$

## Đạo hàm nhiều biến

Đạo hàm riêng (theo từng biến) và đạo hàm toàn phần

## Gradient descent

Phương pháp tối ưu được sử dụng trong ML



### Gradient descent algorithm

repeat until convergence {  
     $\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$   
    (for  $j = 1$  and  $j = 0$ )  
}



# PROBABILITY & STATISTICS

Xác suất và công thức của bayes

Các công thức thống kê, hàm mật độ xác suất



## Xác suất – Định lý bayes

### Sự kiện và xác suất

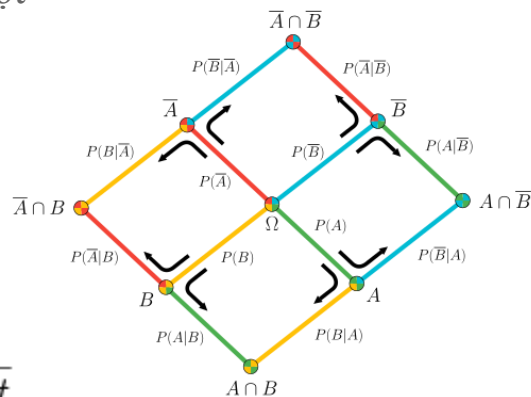
Một sự kiện xảy ra với một xác suất nào đó.

Sự kiện tất yếu, sự kiện bất khả, sự kiện ngẫu nhiên

### Định lý bayes

Xác suất để xảy ra một sự kiện có điều kiện

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\text{likelihood} * \text{prior}}{\text{normalizing\_constant}}$$



$$P(A|B) \cdot P(B) = P(A \cap B) = P(B|A) \cdot P(A)$$



# Thống kê – Hàm mật độ xác suất

## Kỳ vọng

Là giá trị trung bình có  
“trọng số”

$$E[X] = \sum_i p_i x_i$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx.$$

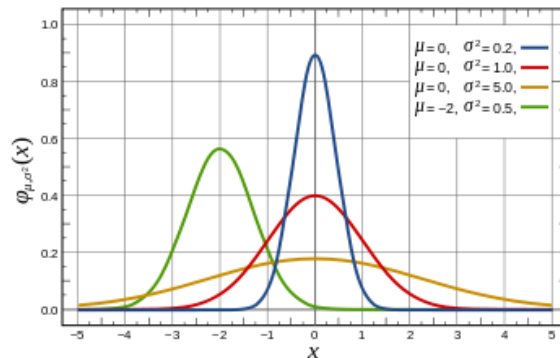
$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

## Hàm mật độ xác suất

Biểu diễn phân bố xác suất  
theo tích phân

Phân bố Gauss (pb chuẩn)

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad f(x | \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



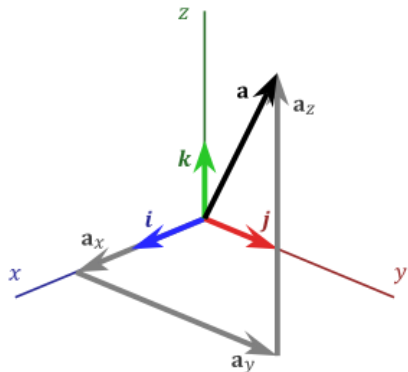




# LET'S REVIEW SOME CONCEPTS

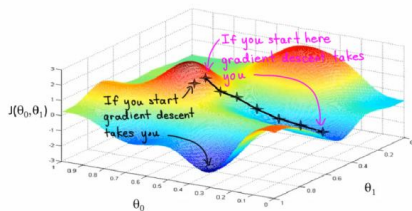
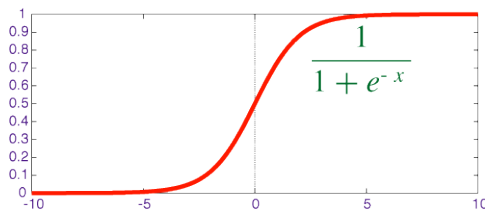
## Linear algebra

Không gian vector  
Scalar, vector, matrix, tensor



## Mathematical Analysis

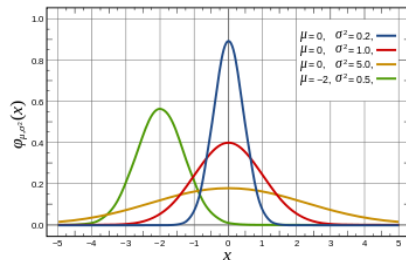
Các hàm số thường gặp  
Phép tính đạo hàm  
Gradient descent



## Statistics

Công thức Bayes  
Các công thức thống kê

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\text{likelihood} * \text{prior}}{\text{normalizing\_constant}}$$





## EXTRACT TASK

Hãy sử dụng  
python hoặc một  
ngôn ngữ bạn biết  
để thực hiện các  
mini tasks đã cho

### Gradient Descent

repeat until convergence:  $\sum$

$$\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)}$$

Simultaneously update  
 $\theta_j$  for all  $j \in \{0, \dots, n\}$

$$\frac{\partial}{\partial \theta_j} J(\theta)$$

}



# THANKS!

## Any questions?

You can find me at  
[caothanhha9@yahoo.com/gmail.com](mailto:caothanhha9@yahoo.com/gmail.com)