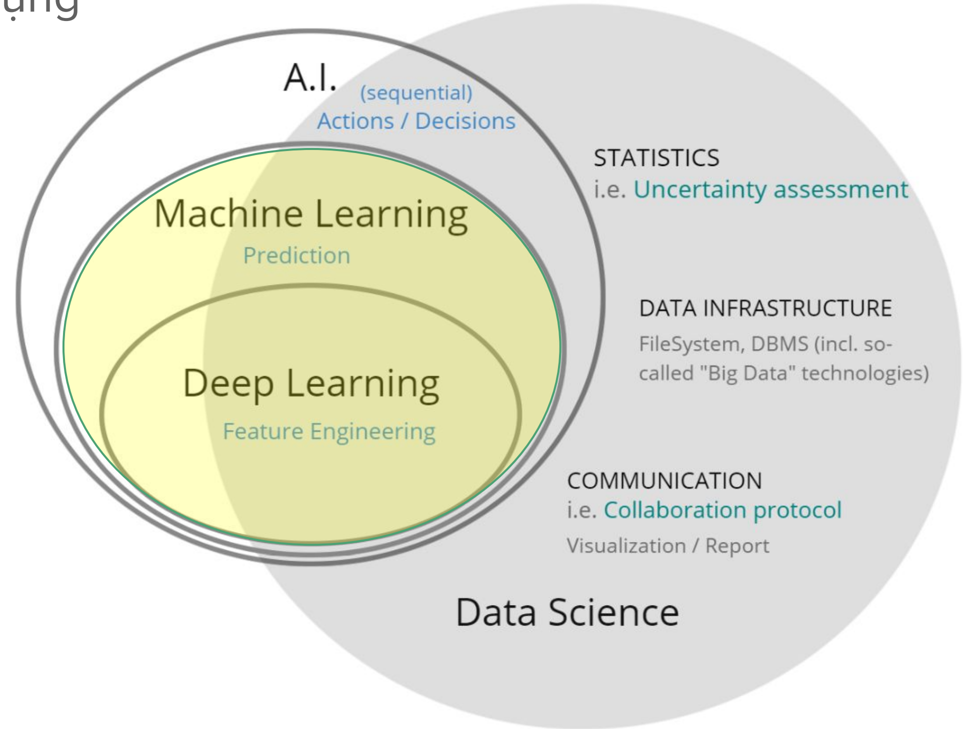
VietAI

Bài 1: Giới thiệu khóa học & Ôn tập kiến thức Toán

Giới thiệu khóa học

Cung cấp kiến thức nền tảng về Machine Learning (ML), Deep Learning (DL) và Giới thiệu một số ứng dụng



A.I. is about *automation*; ML is (mostly) about *automated* prediction with high accuracy* from data;
DL is (mostly) about *automated* feature engineering from data

Nội dung khóa học

NỀN TẢNG MACHINE LEARNING

- **Tuần 1:** Giới thiệu khóa học; Ôn tập kiến thức Toán (linear algebra, probability, calculus); Lập trình Python & NumPy
- **Tuần 2:** Giới thiệu tổng quan về Machine learning và Ứng dụng trong thực tế; Giới thiệu thư viện Tensorflow; Linear Regression và Gradient Descent
- **Tuần 3:** Logistic/Softmax Regression; Overfitting và Validation

Nội dung khóa học

NỀN TẢNG DEEP LEARNING 1 & ỨNG DỤNG TRONG XỬ LÝ ẢNH

- **Tuần 4:** Tổng quan về Deep Neural Network (DNN); Các kỹ thuật quan trọng để huấn luyện DNNs: Thuật toán lan truyền ngược (Backpropagation), Stochastic Gradient Descent, Dropout
- **Tuần 5:** Giới thiệu Convolutional Neural Network (CNN); Các kiến trúc CNN hiện đại

Nội dung khóa học

NỀN TẢNG DEEP LEARNING 2 & ỨNG DỤNG TRONG XỬ LÝ NGÔN NGỮ TỰ NHIÊN

- **Tuần 6:** Ứng dụng của DNN trong Xử lý ngôn ngữ tự nhiên; Giới thiệu về các phương pháp biểu diễn từ (word embeddings) và Language modeling; Giới thiệu Recurrent Neural Network (RNN)
- **Tuần 7:** RNN với kiến trúc Long Short-Term Memory (LSTM) và Gated Recurrent Unit (GRU); Giới thiệu mô hình Sequence-to-Sequence (seq2seq) với kiến trúc encoder-decoder
- **Tuần 8:** Giới thiệu cơ chế Attention; Cơ chế Attention cho mô hình seq2seq; Một số cải tiến của mô hình seq2seq

Nội dung khóa học

CÁC HƯỚNG ĐI TIẾP THEO & TỔNG KẾT

- **Tuần 9:** Một số chủ đề mở & nâng cao *(Guest lectures)*
- **Tuần 10:** Ôn tập, tổng kết & Thuyết trình đồ án khóa học.

Vài lời khuyên nhỏ*

“What-How-Why” pattern

Khi muốn hiểu rõ một vấn đề nào đó, hãy đặt các câu hỏi

1. What? Why (not) that?
2. How? Why (not) so?
3. What else to expect?

“What-How-Why” pattern

(nói riêng cho việc học / bổ sung kiến thức Toán)

“Mathematics is the art of giving the **same name** to **different things**.”

- Henri Poincaré

“The purpose of computation* is **insight**, *not* numbers.”

- Richard Hamming

(quote credit: /c/3blue1brown)

VietAI

Read, Do, Collab!

- Chỉ đọc (Read) mà *không* thực sự làm (Do) vd. không cầm bút suy nghĩ hướng tiếp cận cho một vấn đề gần tương tự, hay không tự tay triển khai một thuật toán liên quan, ... sẽ *rất* dễ dẫn đến việc **ngộ nhận** rằng bản thân đã hiểu *rõ* vấn đề.
- **Cộng tác - Collaboration**:
  - Làm việc nhóm sẽ giúp chúng ta nhìn thấy được nhiều góc cạnh khác nhau của cùng một vấn đề
  - Trong quá trình trao đổi / giảng giải lại một khái niệm hay vấn đề nào đó cho (các) học viên khác, bản thân mỗi người cũng sẽ nhận ra mình đang nắm rõ vấn đề đến mức nào
  - *"Muốn đi nhanh, đi một mình. Muốn đi xa, đi cùng nhau"*

"What I cannot create, I do *not* understand"
- Richard Feynman

(image credit: wikiquote)

Take-home messages

cho phần “Giới thiệu khóa học”

- Kiến thức nền tảng về ML, DL là tiền đề để có thể triển khai các thuật toán AI hiện đại cho ứng dụng hoặc nghiên cứu, nhưng *chưa* phải là toàn thể bức tranh của AI
- Luôn đặt câu hỏi “What-How-Why” khi tiếp cận vấn đề
- Học / Làm việc nhóm sẽ đem lại hiệu quả lâu dài
- *Nội dung tiếp theo:* Kiến thức nền tảng về Toán & Lập trình tính toán Tensor cần thiết để triển khai mô hình / thuật toán ML

VietAI

Ôn tập kiến thức Toán

Đại số tuyến tính, giải tích, lý thuyết xác suất

Đại số tuyến tính (Linear algebra)

Đại số tuyến tính

Outline

- Các khái niệm cơ bản
  - Scalar, Vector, Matrix, Tensor
  - Phép chuyển vị
- Phép toán ma trận
  - Phép cộng ma trận
  - Nhân Vector - Vector, Ma trận - Vector, Ma trận - Ma trận, tích Hadamard
- Ma trận nghịch đảo
  - Ma trận đơn vị
  - Ma trận nghịch đảo

Đại số tuyến tính

Các khái niệm cơ bản

- **Scalar**: $\alpha \in \mathbb{R}$
- **Vector**: $x \in \mathbb{R}^n$ kí hiệu một vector có n phần tử, vector n chiều. Thông thường một vector được viết dưới dạng cột

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Đại số tuyến tính

Các khái niệm cơ bản

- **Ma trận** (**Matrix**): $A \in \mathbb{R}^{m \times n}$ kí hiệu một ma trận có m dòng và n cột

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix}$$

- Ma trận A có được xem như các vector cột (hàng) trong A ghép lại:

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}$$

Đại số tuyến tính

Các khái niệm cơ bản

- **Chuyển vị** (**Transpose**): Thao tác này biến các vector cột của một ma trận thành dòng hoặc ngược lại.

$$A = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_n \\ | & | & & | \end{bmatrix}$$

$$A^\top = \begin{bmatrix} — & a_1^\top & — \\ — & a_2^\top & — \\ & \vdots & \\ — & a_n^\top & — \end{bmatrix}$$

Đại số tuyến tính

Các khái niệm cơ bản

- **Norm**: Norm của một vector cho ta biết "chiều dài" của vector đó. Công thức **Euclidean norm** hay **L2 norm**:

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

- Công thức cho **L1 norm**:

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$

Đại số tuyến tính

Các khái niệm cơ bản

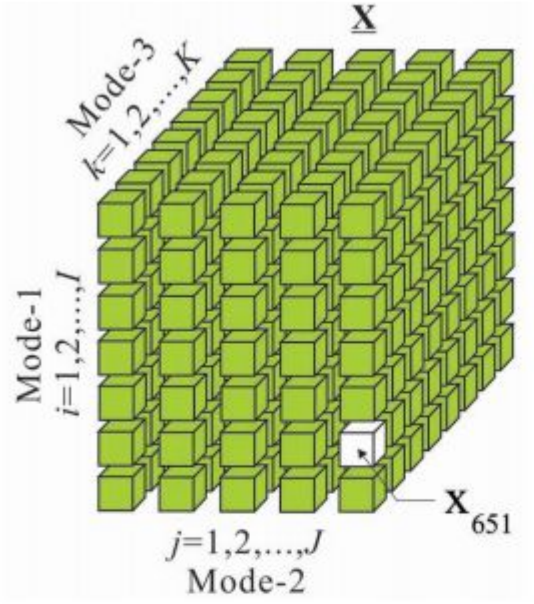
- Công thức L2-norm có thể áp dụng cho ma trận A có kích thước bất kì, khi đó được gọi là **Frobenius norm**

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}^2}$$

Đại số tuyến tính

Các khái niệm cơ bản

- **Tensor**: bao gồm nhiều phần tử được bố trí vào một không gian nhiều hơn 2 chiều. Ví dụ $X \in \mathbb{R}^{m \times n \times p}$ kí hiệu tensor 3 chiều X



Đại số tuyến tính

Phép toán ma trận - Phép cộng ma trận

- Cho hai ma trận có cùng kích thước $A, B \in \mathbb{R}^{m \times n}$ tổng của hai ma trận $C = A + B$

$$A+B = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1n} \\ B_{21} & A_{22} & \dots & B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ B_{m1} & B_{m2} & \dots & B_{mn} \end{bmatrix} = \begin{bmatrix} A_{11}+B_{11} & A_{12}+B_{12} & \dots & A_{1n}+B_{1n} \\ A_{21}+B_{21} & A_{22}+B_{22} & \dots & A_{2n}+B_{21} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}+B_{m1} & A_{m2}+B_{m2} & \dots & A_{mn}+B_{mn} \end{bmatrix}$$

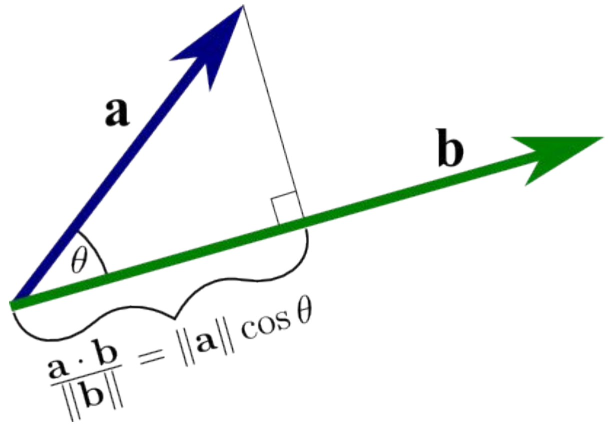
Đại số tuyến tính

Phép toán ma trận - Tích Vector-Vector

- Cho 2 vector $x, y \in \mathbb{R}^n$ kết quả của phép tính sau được gọi là **tích vô hướng** (**inner product** hay **dot product**) của 2 vector x và y:

$$x^\top y = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Chú ý: $x^\top y = y^\top x$



Đại số tuyến tính

Phép toán ma trận - Tích Ma trận - Vector

- Cho ma trận $A \in \mathbb{R}^{m \times n}$ và vector $x \in \mathbb{R}^{n}$, tích của chúng là vector $y = Ax \in \mathbb{R}^{m}$

$$y = Ax = \begin{bmatrix} — & a_1^\top & — \\ — & a_2^\top & — \\ & \vdots & \\ — & a_m^\top & — \end{bmatrix} x = \begin{bmatrix} a_1^\top x \\ a_2^\top x \\ \vdots \\ a_m^\top x \end{bmatrix}$$

Đại số tuyến tính

Phép toán ma trận - Tích Ma trận - Ma trận

- Cho 2 ma trận $A \in \mathbb{R}^{m \times n}$ và $B \in \mathbb{R}^{n \times p}$ Ma trận $C = AB$ là tích của 2 ma trận A và B với $C \in \mathbb{R}^{m \times p}$

$$C_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

$$C = AB = \begin{bmatrix} — & a_1^\top & — \\ — & a_2^\top & — \\ & \vdots & \\ — & a_m^\top & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} a_1^\top b_1 & a_1^\top b_2 & \cdots & a_1^\top b_p \\ a_2^\top b_1 & a_2^\top b_2 & \cdots & a_2^\top b_p \\ \vdots & \vdots & \ddots & \vdots \\ a_m^\top b_1 & a_m^\top b_2 & \cdots & a_2^\top b_p \end{bmatrix}$$

Đại số tuyến tính

Phép toán ma trận - Tính chất

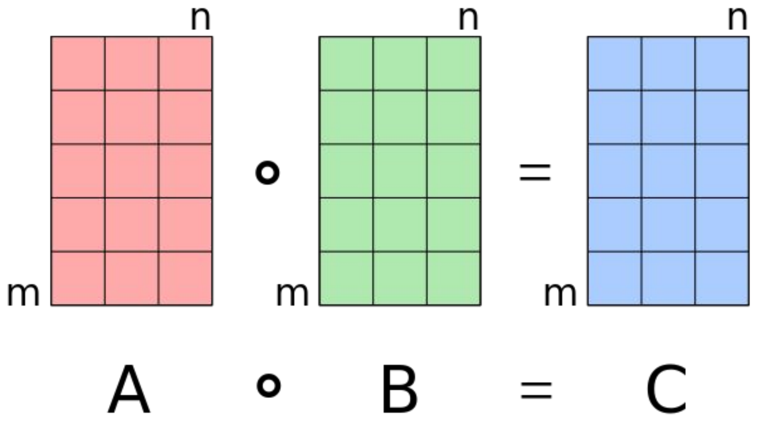
- Phép nhân ma trận **không có** tính giao hoán $AB \neq BA$
- Kết hợp: $(AB)C = A(BC)$
- Phân phối: $A(B+C) = AB + AC$
- Transpose:
  - $(A^\top)^\top = A$
  - $(AB)^\top = B^\top A^\top$
  - $(A+B)^\top = A^\top + B^\top$

Đại số tuyến tính

Phép toán ma trận - Tích Hadamard

- Cho hai ma trận có cùng kích thước $A, B \in \mathbb{R}^{m \times n}$, **tích Hadamard** (***entry-wise product***) . $C = A \circ B$

$$C_{ij} = A_{ij} B_{ij}$$



Đại số tuyến tính

Ma trận nghịch đảo

- **Ma trận đơn vị** (***Indentity Matrix***) kí hiệu là $I_n \in \mathbb{R}^{n \times n}$ là ma trận vuông với tất cả phần tử trên đường chéo chính bằng 1, các phần tử còn lại bằng 0

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} \qquad I_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Với mọi ma trận $A \in \mathbb{R}^{m \times n}$:

$$AI_n = A = I_m A$$

Đại số tuyến tính

Ma trận nghịch đảo

- **Ma trận nghịch đảo** (***Inverse Matrix***) của ma trận $A \in \mathbb{R}^{n \times n}$ kí hiệu là $A^{-1} \in \mathbb{R}^{n \times n}$ ma trận sao cho:

$$A^{-1}A = I_n = AA^{-1}$$

- Một ma trận vuông có thể không có ma trận nghịch đảo khi đó nó được gọi là ***non-invertible*** hay ***singular***. Nếu nó có thì chỉ có duy nhất một ma trận nghịch đảo.

Giải tích (Calculus)

Giải tích

Outline

- Các công thức đạo hàm cơ bản
- Đạo hàm riêng
- Gradient

Giải tích

Các công thức cơ bản

Đạo hàm (***Derivatives***) là khái niệm quan trọng trong giải tích để tìm giá trị lớn nhất (hoặc nhỏ nhất) của hàm số

- $(x^n)' = nx^{n-1}$
- $(e^x)' = e^x$
- $(\log x)' = \frac{1}{x}$
- $\left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$
- $(uv)' = u'v + v'u$
- $[f(g(x))]' = f'(g(x))g'(x)$

Giải tích

Đạo hàm riêng

- **Đạo hàm riêng** (***Partial Derivatives***) của hàm nhiều biến f theo biến x kí hiệu là $\frac{\partial f}{\partial x}$, để hiểu được khái niệm này ta xét ví dụ hàm $f\left(x,y\right)$ sau:

$$f\left(x,y\right)=x^2+y^2-3xy-x+e^x+e^y$$

$$\frac{\partial f}{\partial x}=2x-3y-1+e^x$$

$$\frac{\partial f}{\partial y}=2y-3x+e^y$$

Giải tích

Gradient

- Cho hàm số $f : \mathbb{R}^n \to \mathbb{R}$ gradient của f là vector có n phần tử bao gồm các đạo hàm riêng của f :

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

Lý thuyết xác suất (Probability Theory)

Lý thuyết xác suất

Outline

- Các khái niệm cơ bản trong xác suất
- Xác suất có điều kiện
- Sự độc lập của các sự kiện
- Biến ngẫu nhiên
- Phân phối đều
- Phân phối chuẩn

Lý thuyết xác suất

Khái niệm cơ bản

- **Không gian mẫu** (***Sample space***): Thường được kí hiệu là Ω , là tập hợp bao gồm tất cả các kết quả có thể xảy ra của một phép thử. Ví dụ tung một viên xúc sắc có 6 mặt, khi đó $\Omega = \{1, 2, 3, 4, 5, 6\}$
- **Sự kiện** (***Event***): Một tập hợp con của Ω
- **Tiên đề xác suất** (***Probability Axioms***)

– $P(A) \geq 0$

– $P(\Omega) = 1$

– Nếu $A \cap B = \varnothing$ thì $P(A \cup B) = P(A) + P(B)$

Lý thuyết xác suất

Khái niệm cơ bản

- **Tính chất:**
  - Nếu $A \subseteq B$ thì $P(A) \leq P(B)$
  - Union bound: $P(A \cup B) \leq P(A) + P(B)$
  - Nếu $A_1, \ldots, A_k$ là các sự kiện không giao nhau từng đôi một và $\cup_{i=1}^{k} A_i = \Omega$, thì $\Sigma_{i=1}^{k} P(A_i) = 1$

Lý thuyết xác suất

Xác suất có điều kiện

- Gọi B là sự kiện có xác suất lớn hơn 0. Xác suất có điều kiện của sự kiện A khi B đã xảy ra ký hiệu $P\left(A|B\right)$được tính theo công thức:

$$P\left(A|B\right) = \frac{P\left(AB\right)}{P\left(B\right)}$$

Lý thuyết xác suất

Xác suất có điều kiện

- Tính chất
  - $P(A|B) \geq 0$
  - $P(\Omega|B) = 1$
  - Nếu $A \cap C = \varnothing$ thì $P(A \cup C|B) = P(A|B) + P(C|B)$
- Công thức nhân xác suất:
  - $P(AB) = P(B)\,P(A|B) = P(A)\,P(B|A)$
  - $P(A_1 A_2 \ldots A_n) = P(A_1) \prod_{i=2}^{n} P(A_i|A_1 \ldots A_{i-1})$
- Công thức Bayes:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Lý thuyết xác suất

Sự độc lập của các sự kiện (Indepedence of events)

- Hai sự kiện A và B được gọi là độc lập với nhau nếu:

$$P(AB) = P(A)\,P(B)$$

- Hệ quả là $P(A|B) = P(A)$hay $P(B|A) = P(B)$ Hay nói cách khác việc A xảy ra không cung cấp thêm thông tin liên quan đến B và ngược lại.

Lý thuyết xác suất

Biến ngẫu nhiên

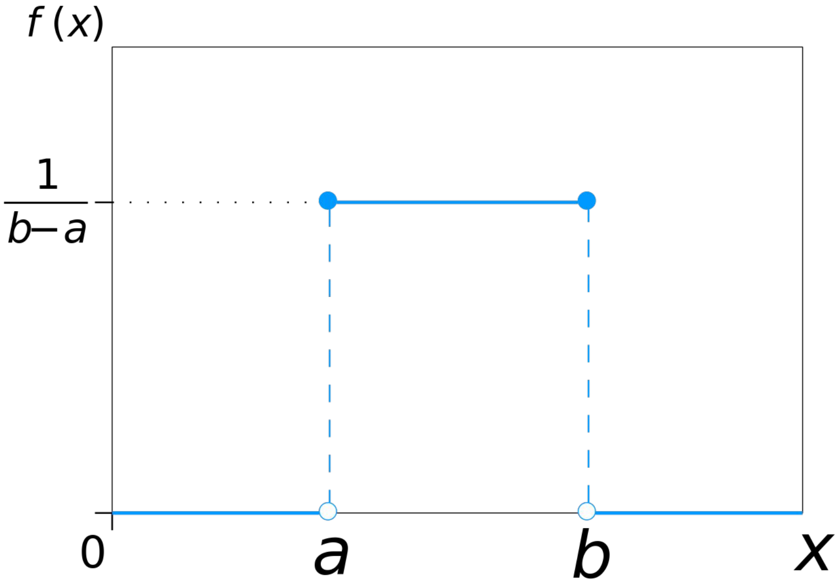
- **Biến ngẫu nhiên** (***random variable***) gán mỗi kết quả của phép thử với một số. Có thể định nghĩa một cách toán học là một hàm số $X : \Omega \to \mathbb{R}$
    - Biến ngẫu nhiên rời rạc.
    - Biến ngẫu nhiên liên tục.

Lý thuyết xác suất

Biến ngẫu nhiên - Phân phối đều

- $X \sim Uniform\,(a, b)$ Xác suất như nhau tại mọi giá trị trong khoảng từ a đến b

$$f\,(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$



Lý thuyết xác suất

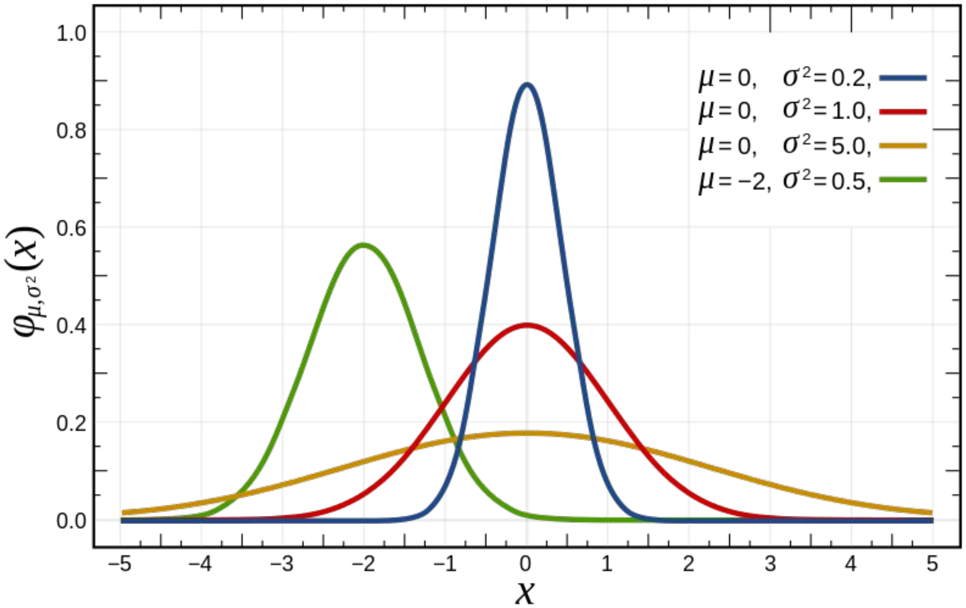
Biến ngẫu nhiên - Phân phối chuẩn

Lý thuyết xác suất

Biến ngẫu nhiên - Phân phối chuẩn

- $X \sim Normal\left(\mu, \sigma^2\right)$Phân phối chuẩn/Gauss (Normal/Gaussian Distribution)

$$f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



VietAI

Bài tập

1. Thực hiện các phép tính sau:

(a) $$\begin{bmatrix} 2 \\ -1 \end{bmatrix}^\top \begin{bmatrix} 3 \\ -2 \end{bmatrix}$$

(b) $$\begin{bmatrix} -1 \\ 2 \\ 3 \end{bmatrix}^\top \begin{bmatrix} 4 \\ -3 \\ 2 \end{bmatrix}$$

(c) $$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

(d) $$\begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

(e) $$\begin{bmatrix} -2 & 3 & 1 \\ 0 & 4 & -1 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

(f) $$\begin{bmatrix} 2 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

(g) $$\begin{bmatrix} 2 & 1 & 2 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 3 & 1 \\ -1 & 1 & 1 \end{bmatrix}$$

VietAI

Bài tập

2. Tính vector gradient của các hàm số sau:

(a) $f(x, y) = x^2 + xy + y^2 + x + 1$ với $x, y \in \mathbb{R}$

(b) $f(z) = w^\top z$ với $w, z \in \mathbb{R}^n$

(c) $f(v) = v^\top v$ với $v \in \mathbb{R}^n$

Tài liệu tham khảo

1. *Linear algebra review* - CS229 - Stanford University
   http://cs229.stanford.edu/section/cs229-linalg.pdf
2. Paul's Online Math Notes - http://tutorial.math.lamar.edu/