Lecture 02
Math Review

Prof. Ph.D. Woo Youn Kim

Contents

- Linear algebra
  - Vector, Matrix, Tensor

- Probability

  http://cs229.stanford.edu/section/cs229-prob.pdf

- Principle Component Analysis

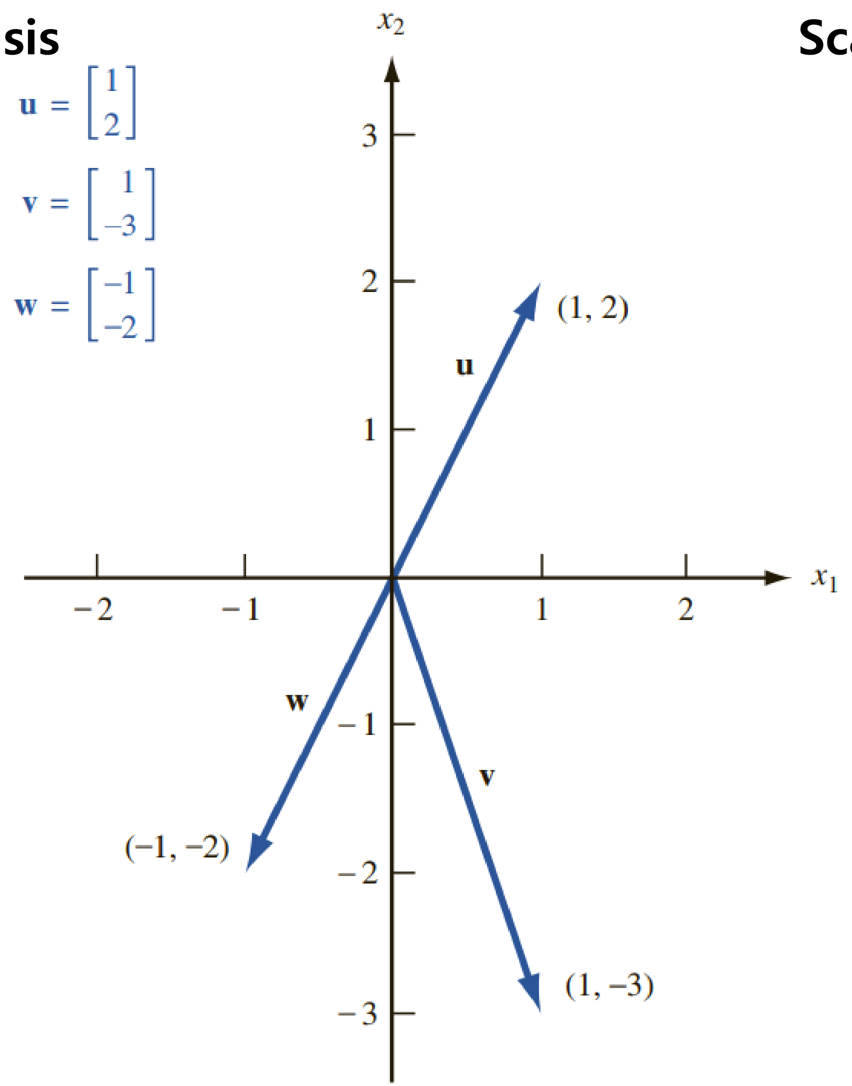
Vector

Vector

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

Basis

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \quad \mathbf{w} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$$



Span

Scalar product of u and v (written $u \cdot v$)

$$u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos \theta,$$

the length of the vector

Θ is the angle between the two vectors

Matrix

Matrix

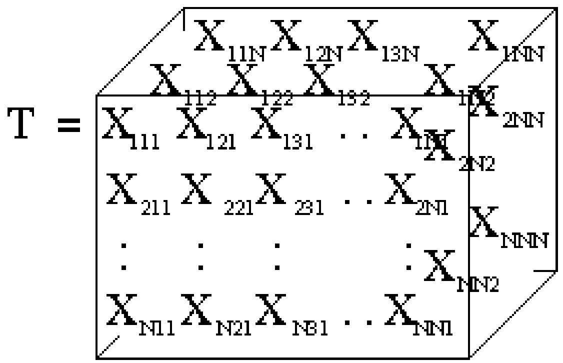
$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

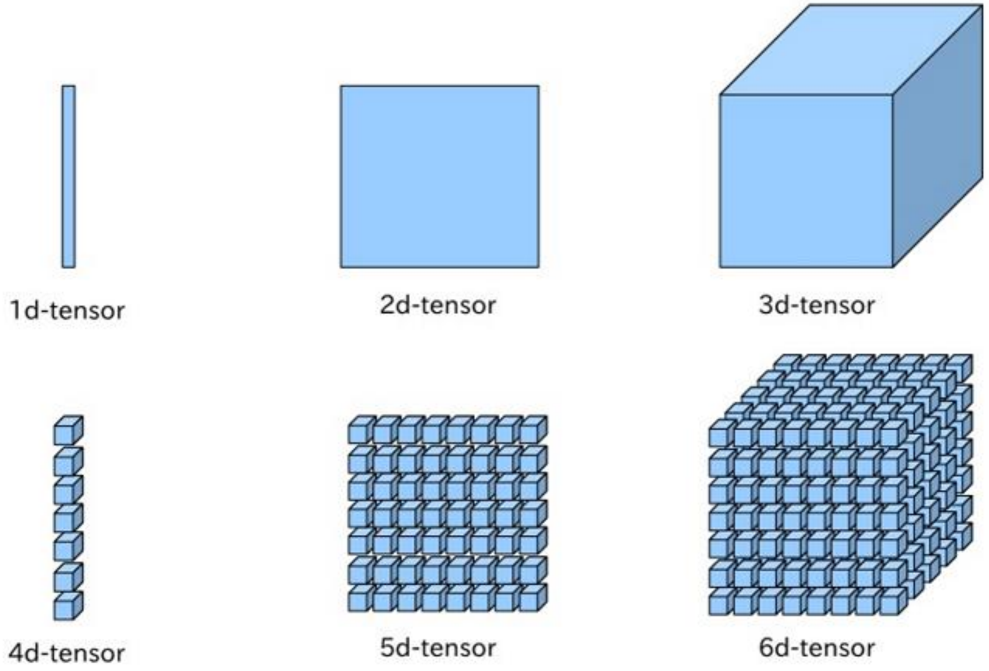
For example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 \end{bmatrix}$$

Tensor

Natural extension of vector and matrix to higher dimension: n-th rank tensor





TensorFlow

Matrix addition

Addition of vectors

$$\text{xA+yB} = x\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} + y\begin{bmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & \ddots & \vdots \\ b_{n1} & \cdots & b_{nm} \end{bmatrix} = \begin{bmatrix} xa_{11} + yb_{11} & \cdots & xa_{1m} + yb_{1m} \\ \vdots & \ddots & \vdots \\ xa_{n1} + ybn_1 & \cdots & xa_{nm} + ybnm \end{bmatrix}$$

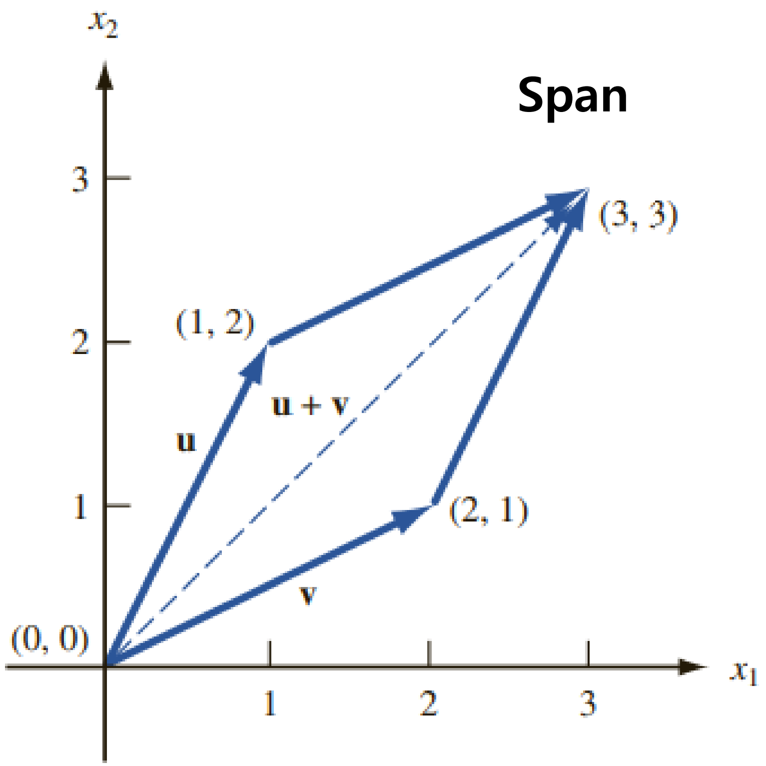
$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1 & -2 & -3 \\ 2 & 1 & -1 \end{bmatrix}$$

$$\mathbf{v} = [2 \;\; 1]$$
$$\mathbf{u} = [1 \;\; 2]$$
$$\mathbf{u} + \mathbf{v} = [3 \;\; 3]$$

then

$$A + B = \begin{bmatrix} 1-1 & 2-2 & 3-3 \\ 0+2 & -1+1 & 1-1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \end{bmatrix}.$$

Span



Transpose of matrix

Transpose

Given any $m \times n$ matrix

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

the **transpose** of A (written A^T) is the $n \times m$ matrix

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

Matrix multiplication

Multiplication

The **matrix product** $C = AB$ of A and B is the $m \times n$ matrix C whose ijth element is determined as follows:

ijth element of C = scalar product of row i of A $\times$ column j of B ■

$$c_{ij} = \sum_{m=1}^{k} a_{im} b_{mj}$$

Number of columns in A = number of rows in B

Compute $C = AB$ for

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 1 & 2 \end{bmatrix}$$

$$C = AB = \begin{bmatrix} 5 & 8 \\ 7 & 11 \end{bmatrix}$$

Matrix multiplication

Matrix multiplication is associative. A(BC) = (AB)C

$$d_{ik} = \sum_{j=1}^{N} a_{ij}\left(\sum_{n=1}^{N} b_{jn}c_{nk}\right) = \sum_{j=1}^{N}\sum_{n=1}^{N} a_{ij}b_{jn}c_{nk} = \sum_{j=1}^{N}\sum_{n=1}^{N}(a_{ij}b_{jn})c_{nk} = \sum_{n=1}^{N}\left(\sum_{j=1}^{N} a_{ij}b_{jn}\right)c_{nk}$$

Matrix multiplication is distributive. A(B + C) = AB + AC

Inverse of matrix

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

For A to have its inverse, the determinant of A must be non-zero.

Example

n number of simultaneous equations

$$A_{11}x_1 + A_{12}x_2 + ... + A_{1n}x_n = b_1$$

$$A_{21}x_1 + A_{22}x_2 + ... + A_{2n}x_n = b_2$$

$$A_{n1}x_1 + A_{m2}x_2 + ... + A_{n,n}x_n = b_n$$

$$\boldsymbol{Ax=b}$$

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

Example

n number of simultaneous equations

$$\boldsymbol{Ax}=\boldsymbol{b} \qquad A=\begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \vdots & \vdots \\ A_{n,1} & \cdots & A_{nn} \end{bmatrix} \qquad \boldsymbol{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad \boldsymbol{b}=\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

If A is invertible,

$$A\boldsymbol{x}=\boldsymbol{b}$$

$$A^{-1}A\boldsymbol{x}=A^{-1}\boldsymbol{b}$$

$$I_n\boldsymbol{x}=A^{-1}\boldsymbol{b}$$

$$\boldsymbol{x}=A^{-1}\boldsymbol{b}$$

Norms

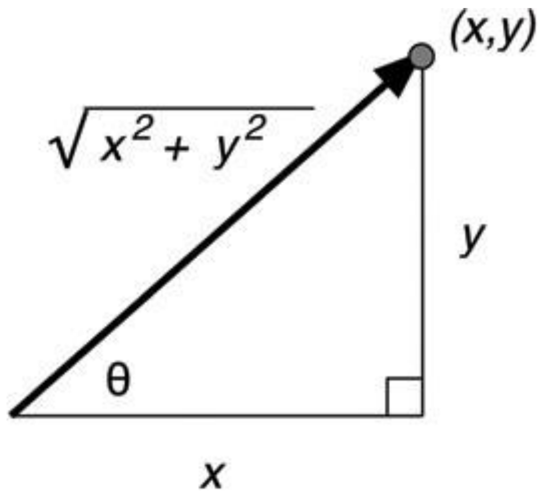
Used to measure the size of a vector

Norm of a vector **x** is distance from origin to x

Definition of L^P norm

$$\|x\|_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$

- L^2 Norm
  - called Euclidean norm, written simply as $\|x\|$
  - squared Euclidean norm is same as $x^T x$



Question

- L_∞ Norm

$$\|\boldsymbol{x}\|_\infty = \max_i |x_i|$$

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$$

  - called max norm

Justify the above formula using the definition of L^P norm.

Probability

- Joint probability

  $P(\mathrm{A,B}) = P(\mathrm{A \cap B})$ probability measure of the simultaneous events A and B

- Conditional probability of any event A given B

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$$

✓ P(A|B) is the probability measure of the event A after observing the occurrence of event B.

✓ For two independent events A and B,

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

➔ $P(A|B) = P(A)$

KAIST

Product rule

$$P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}$$

➡ $P(A,B) = P(A|B)P(B)$

➡ $P(B,A) = P(B|A)P(A)$

Because $P(A,B) = P(B,A)$,

$$P(A|B)P(B) = P(B|A)P(A)$$

Question

60% of students pass the final and 30% of students pass both the final and the midterm

What percent of students who passed the final also passed the midterm?

$P(M|F) = P(M,F) / P(F) = 0.3/0.6 = 0.5$

Marginalization

- Marginalization (Sum Rule)

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

- Law of total probability

$$P(A) = \sum_B P(A,B) = \sum_B P(A|B)P(B)$$

⇐ $P(A,B) = P(A|B)P(B)$

Bayes' rule

$P(A|B) = P(A,B)/P(B)$ using the conditional probability

$= P(B|A)P(A)/P(B)$ using the product rule

$= P(B|A)P(A)/\sum_A P(B|A)P(A)$ using the law of total probability

$$\Rightarrow \quad P(A|B) = \frac{P(B|A)P(A)}{\sum_A P(B|A)P(A)}$$

Question

- Suppose you have tested positive for a disease; what is the probability that you actually have the disease?

- It depends on the accuracy and sensitivity of the test, and on the background (prior) probability of the disease.

$P(D=1) = 0.01$ (background or prior probability)

$P(T=1|D=1) = 0.95$ (true positive)

$P(T=1|D=0) = 0.10$ (false positive)

data we already secured

$P(D=1|T=1) = ?$ posterior probability

prediction based on the data

Question

$P(T=1|D=1) = 0.95$ (true positive)

$P(T=1|D=0) = 0.10$ (false positive)

$P(D=1) = 0.01$ (background or prior probability)

$P(D=1|T=1) = ?$

Bayes' Rule

$$P(D=1|T=1) = P(T=1|D=1)P(D=1) \,/\, P(T=1)$$
$$= 0.95*0.01/\, P(T=1)$$
$$= 0.95*0.01/0.1085 = 0.087$$

$$P(T=1) = \sum_D P(T|D)P(D)$$
$$= P(T=1|D=1)P(D=1)+P(T=1|D=0)P(D=0)$$
$$= 0.95*0.01 + 0.1*0.99 = 0.1085$$

The probability that you have the disease given you tested positive is 8.7%

Likelihood

Bayes' Rule

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} = \frac{P(x|\theta)P(\theta)}{\sum_{\theta} P(x|\theta)P(\theta)}$$

$$\text{Posterior probability} = \frac{(\text{likelihood}) \times (\text{prior probability})}{(\text{evidence})}$$

Prior probability (사전확률) $P(\theta)$: probability of a parameter set θ.

Posterior probability (사후확률) $P(\theta|x)$: $P(\theta)$ given an observation x.

Evidence (증거) $P(x)$: probability of the observation.

Likelihood (가능도) $L(\theta|x) = P(x|\theta)$: $P(x)$ given the parameters θ

θ = disease

x = test result (T or F)

$P(x)$: probability of T and F

$P(\theta)$: probability of having a disease

$P(\theta|x)$: probability of having the disease given the test result

$P(x)$: probability of T and F.

Example

Suppose a coin flip with a single parameter θ_{H} that expresses the "fairness" of the coin or probability that the given coin lands heads up (H).

For a perfectly fair coin, $\theta_{\mathrm{H}} = 0.5$

Imagine flipping a coin twice, and observing the following data : two heads in two tosses ("HH").

$$P(HH|\ \theta_{\mathrm{H}} = 0.5) = 0.5^2 = 0.25$$

What is the likelihood of this problem?

HH: observation
θ_{H} : parameter

Thus, $L(\theta_{\mathrm{H}} = 0.5|HH) = P(HH|\ \theta_{\mathrm{H}} = 0.5) = 0.25$

Question

Suppose an imperfect coin with $\theta_H = 0.3$.

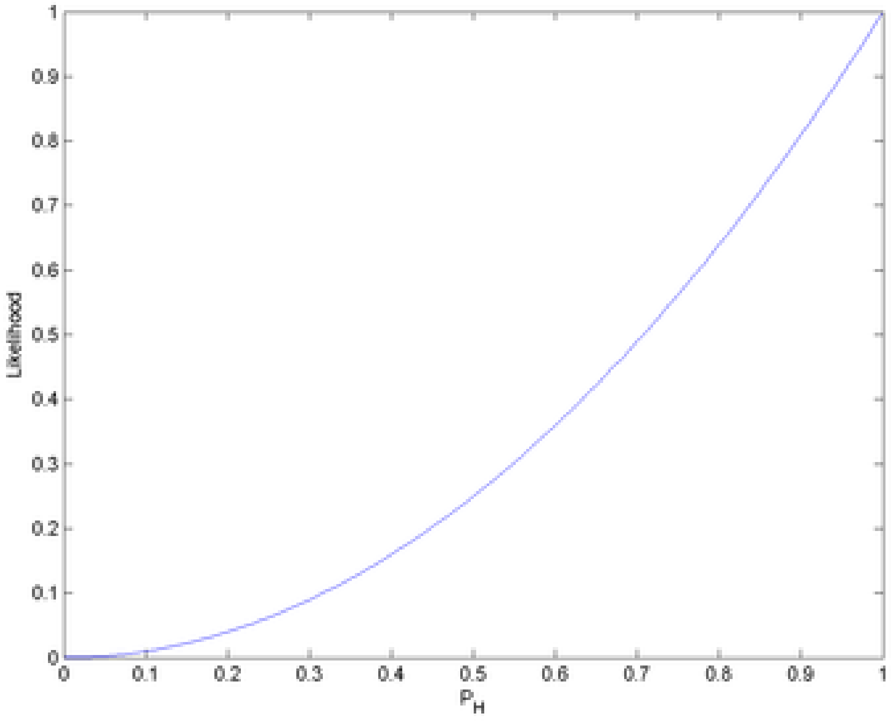
What is the probability of getting two heads?

$P(HH|\ \theta_H = 0.3) = ?$

$P(HH|\ \theta_H = 0.3) = 0.3^2 = 0.09$

What is the likelihood of this problem?

$L(\theta_H = 0.3|HH) = P(HH|\ \theta_H = 0.3) = 0.09$



likelihood function

Note that the integration of the likelihood function is not normalized to 1, meaning that it is not a probability function.