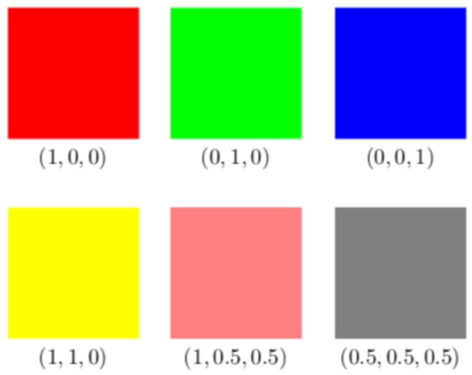
# Big Data and Data Mining

03. Basic Mathematics

# Linear Algebra – Vector and Matrix



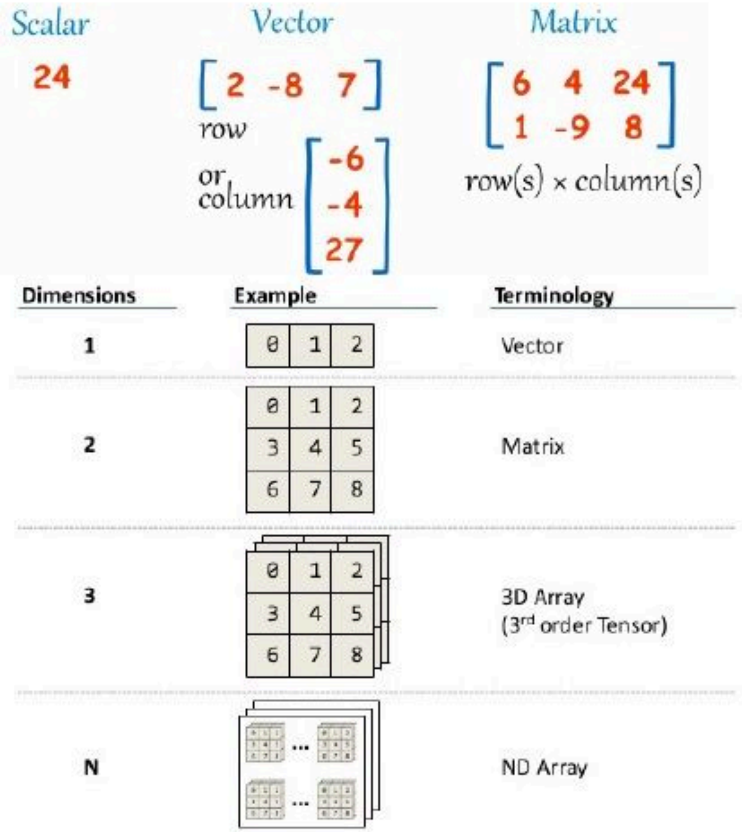

- A vector is an ordered finite list of numbers.
  - Numerical List

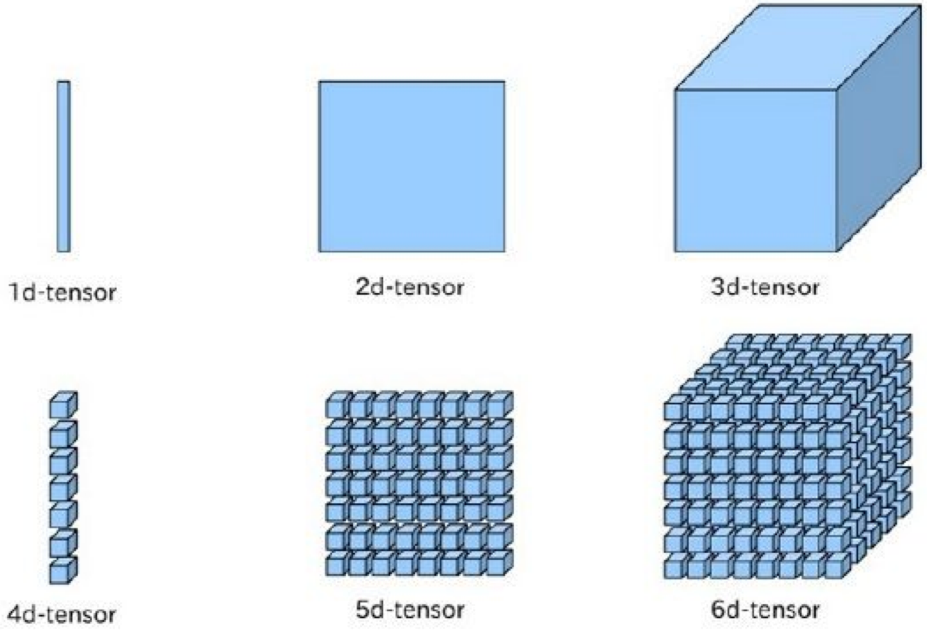
$$\begin{bmatrix} -1.1 \\ 0.0 \\ 3.6 \\ -7.2 \end{bmatrix} \quad \text{or} \quad \begin{pmatrix} -1.1 \\ 0.0 \\ 3.6 \\ -7.2 \end{pmatrix} \quad \text{or} \quad (-1.1, 0.0, 3.6, -7.2)$$

- A matrices is a rectangular array of numbers written between rectangular brackets.

$$\begin{bmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{bmatrix} \quad \text{or} \quad \begin{pmatrix} 0 & 1 & -2.3 & 0.1 \\ 1.3 & 4 & -0.1 & 0 \\ 4.1 & -1 & 0 & 1.7 \end{pmatrix}$$

# Scalar, Vector, Matrices, Tensor




a) https://pic1.zhimg.com/80/v2-c90867d134da775194ee98a97def9ea4_hd.jpg
b) https://pic3.zhimg.com/80/v2-6dd8366ecb7fdb05c65c284d2468321e_hd.jpg

# Norm and Distance

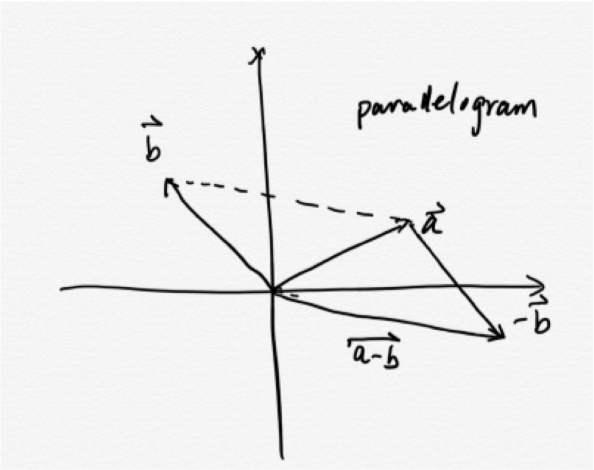
- The Euclidean norm of an n-vector x (named after the Greek mathematician Euclid), denoted ‖x‖, is the square root of the sum of the squares of its elements:

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$$

$$\left\| \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} \right\| = \sqrt{9} = 3, \qquad \left\| \begin{bmatrix} 0 \\ -1 \end{bmatrix} \right\| = 1.$$

- We can use the norm to define the Euclidean distance between two vectors a and b as the norm of their difference:

$$\mathbf{dist}(a, b) = \|a - b\|.$$

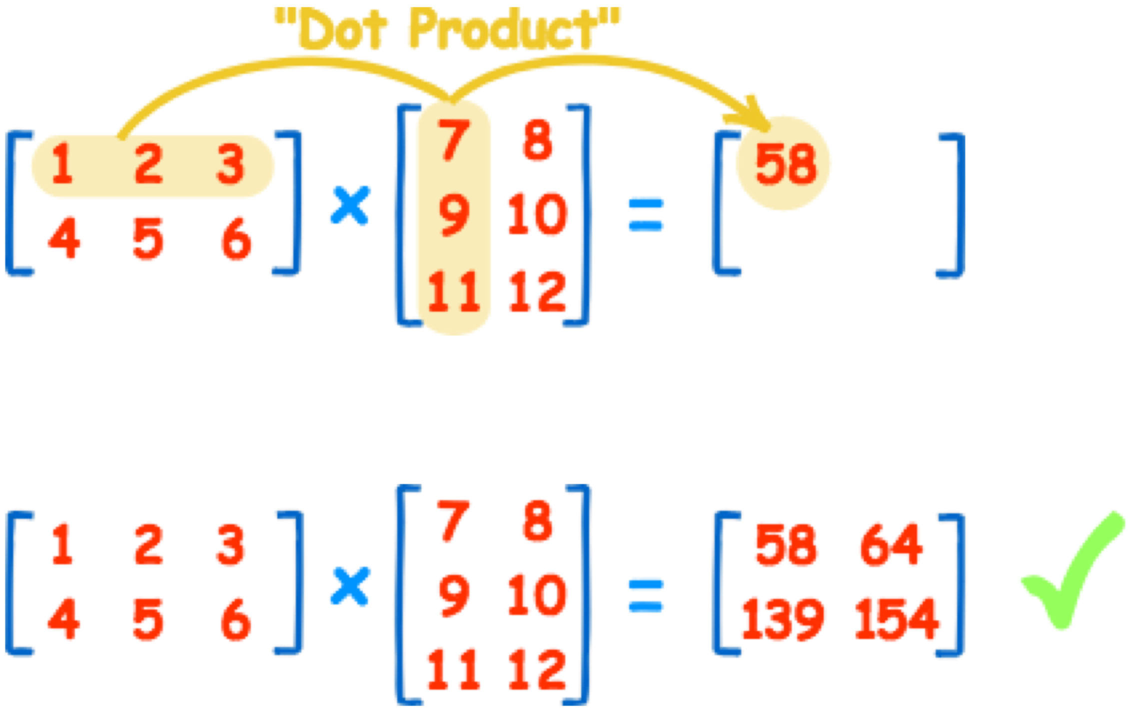


# Linear Equation

$$\begin{cases} 2x - y = 0 \\ -x + 2y = 3 \end{cases}$$

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\boldsymbol{A}\,\boldsymbol{X} = \boldsymbol{B}$$



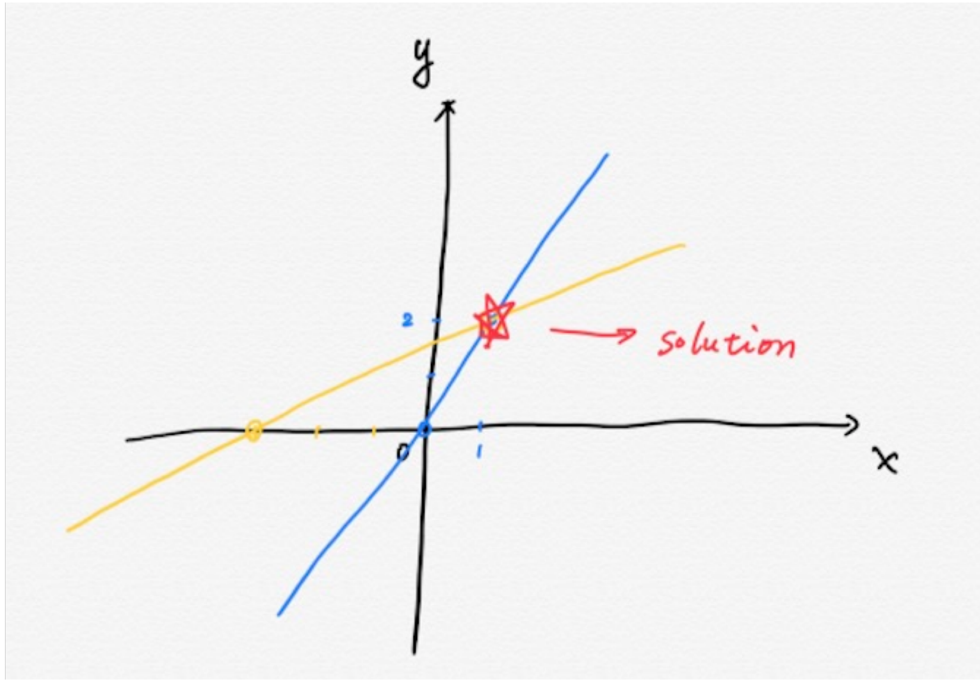

a) https://www.mathsisfun.com/algebra/matrix-multiplying.html

# Matrix – Row Picture

- The key concept of row picture is so-called equation solving.

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\begin{cases} 2x - y = 0 \\ -x + 2y = 3 \end{cases}$$

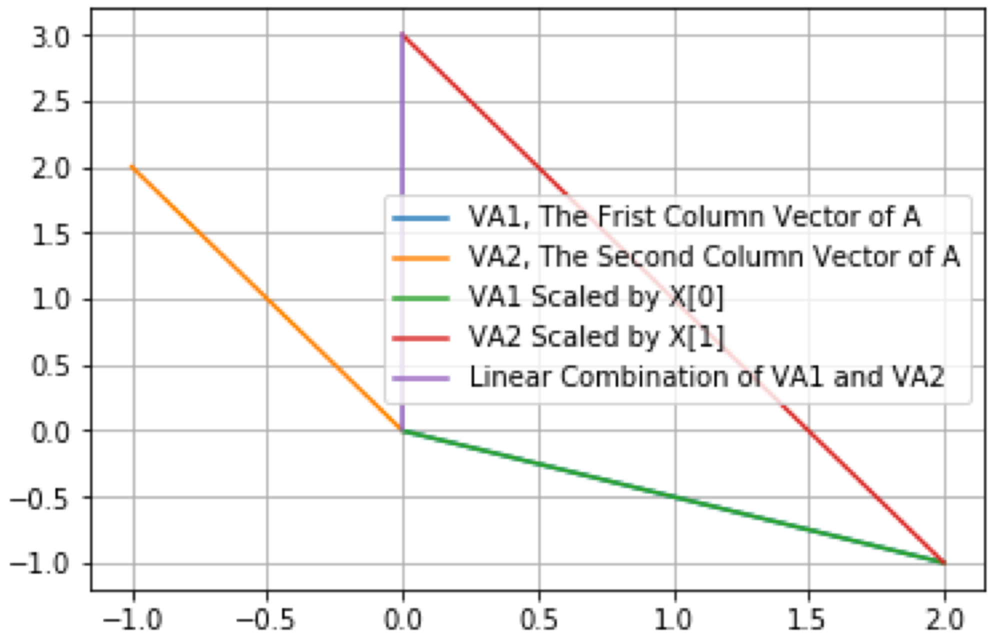


# Matrix – Column Picture

- The key concept of column picture is so-called linear combination.

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ -1 \end{bmatrix} x + \begin{bmatrix} -1 \\ 2 \end{bmatrix} y = \begin{bmatrix} 0 \\ 3 \end{bmatrix}$$



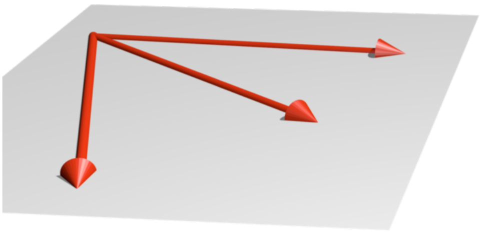
# Linear Independent

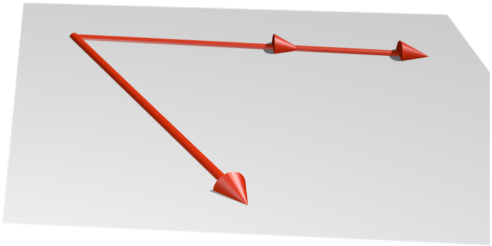
- **Linear Dependent:** A collection or list of n-vectors $a_1, \ldots, a_k$ (with $k \geq 1$) is called linearly dependent if

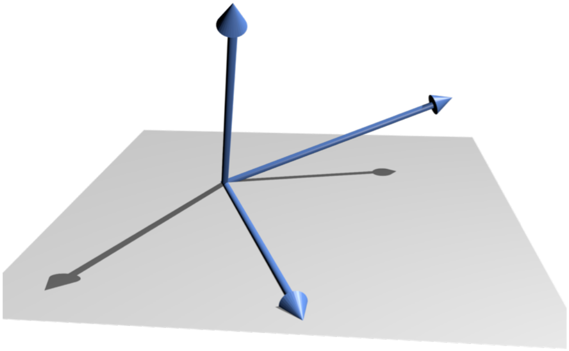
$$\beta_1 a_1 + \cdots + \beta_k a_k = 0$$

holds for some $\beta_1,\ldots,\beta_k$ that are not all zero.

- **Linear Independent:** A collection of n-vectors $a_1, \ldots, a_k$ (with $k \geq 1$) is called linearly independent if it is not linearly dependent, which means that only holds for $\beta_1 = \cdots = \beta_k = 0$.







# Basis for Vector Space

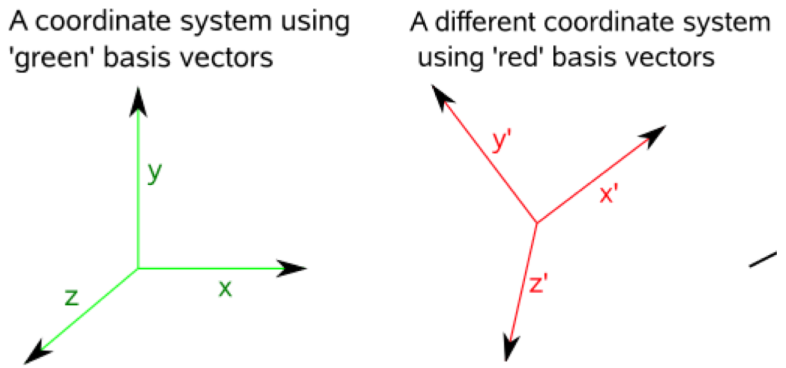
- **Linear combinations of linearly independent vectors.** Suppose a vector x is a linear combination of $a_1, \dots, a_k$,

$$x = \beta_1 a_1 + \cdots + \beta_k a_k.$$

- When the vectors $a_1, \dots, a_k$ are linearly independent, the coefficients that form x are unique: If we also have

$$x = \gamma_1 a_1 + \cdots + \gamma_k a_k,$$

- then $\beta_i = \gamma_i$ for $i = 1,\dots,k$.




a) http://www.euclideanspace.com/maths/algebra/matrix/tensor/coordinates/index.htm

# Matrix Inverse and Decomposition

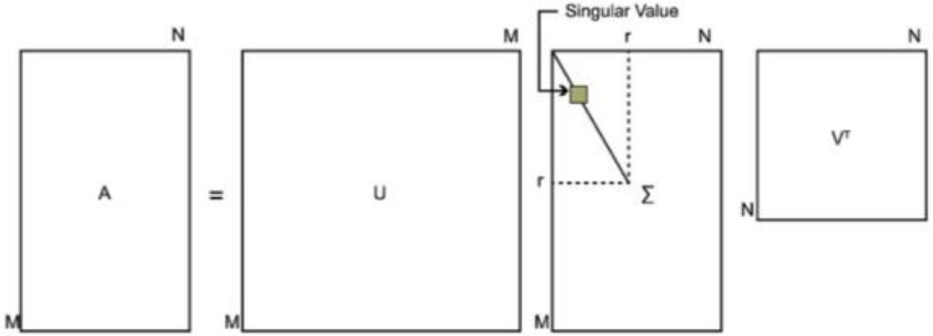
- Solving Linear Equation with matrix inverse

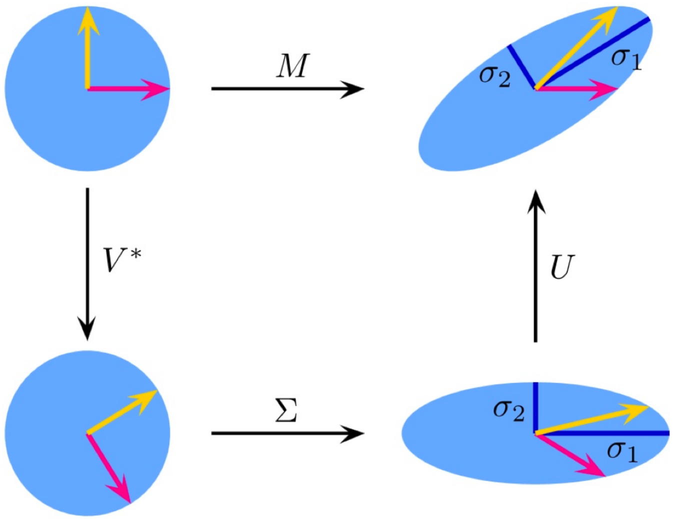
$$\boldsymbol{A}\,{\color{red}\boldsymbol{X}} = \boldsymbol{B}$$
$$\boldsymbol{A}^{-1}\boldsymbol{A}\,{\color{red}\boldsymbol{X}} = \boldsymbol{A}^{-1}\boldsymbol{B}$$
$${\color{red}\boldsymbol{X}} = \boldsymbol{A}^{-1}\boldsymbol{B}$$

- Singular Value Decomposition (SVD)

$$\boldsymbol{A} = \boldsymbol{U\Sigma V}^{*}$$ U and V* are unitary matrices

$$\boldsymbol{A}^{-1} = \boldsymbol{V\Sigma}^{-1}\boldsymbol{U}^{*}$$



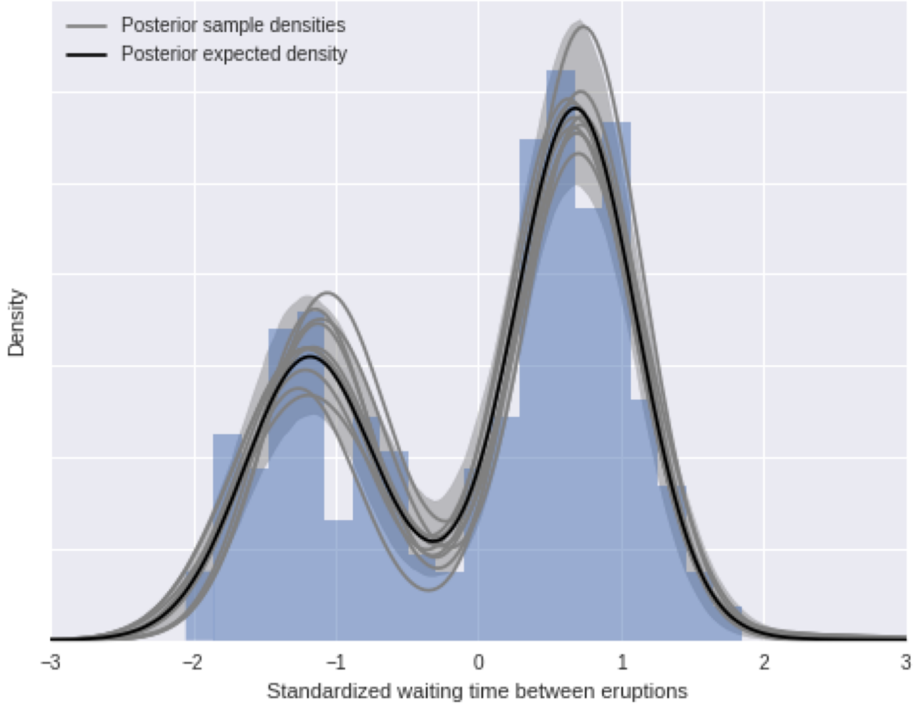



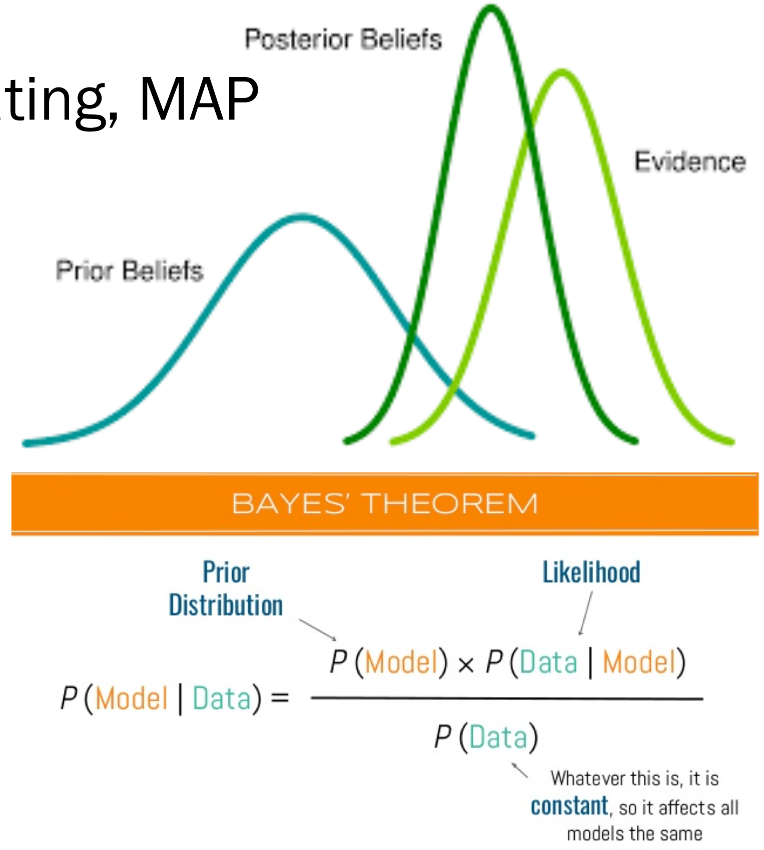

$$M = U \cdot \Sigma \cdot V^{*}$$

a) https://en.wikipedia.org/wiki/Singular_value_decomposition

# Probability and Statistics

- Probability is the measure of the likelihood that an event will occur .
  - Frequentist Statistics, Counting, MLE
  - Bayesian Statistics, Observing and Updating, MAP




a) https://www.analyticsvidhya.com/blog/2016/06/bayesian-statistics-beginners-simple-english/
b) https://www.slideshare.net/andreslopezsepulcre/foundations-of-statistics-in-ecology-and-evolution-8-bayesian-statistics

# Conditional Probability – Bayesian

| | Number of times students visited tutoring | | | |
|---|---|---|---|---|
| | One or fewer times | Two to three times | Four or more times | Total |
| Full time student | 12 | 25 | 8 | 45 |
| Part time student | 2 | 5 | 6 | 13 |
| Total | 14 | 30 | 14 | 58 |

P(four or more times| full time student) $= \frac{8}{45} \approx \boxed{0.18}$

find given

| | Number of times students visited tutoring | | | |
|---|---|---|---|---|
| | One or fewer times | Two to three times | Four or more times | Total |
| Full time student | 12 | 25 | 8 | 45 |
| Part time student | 2 | 5 | 6 | 13 |
| Total | 14 | 30 | 14 | 58 |

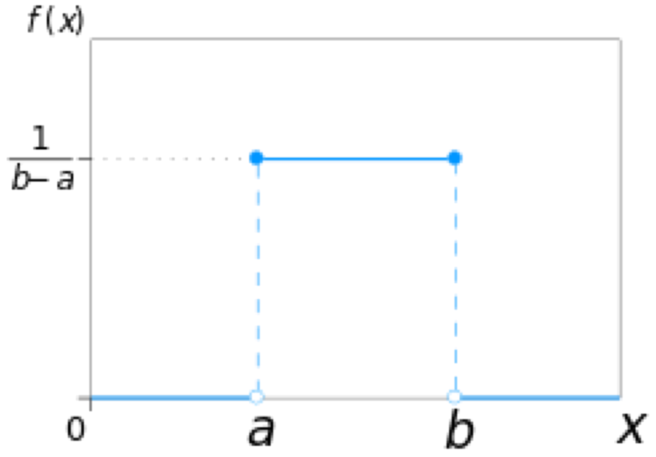
P( part time | visited four or more times) $= \frac{6}{14} \approx \boxed{0.43}$

find given

a) https://www.mathbootcamps.com/conditional-probability-notation-calculation/

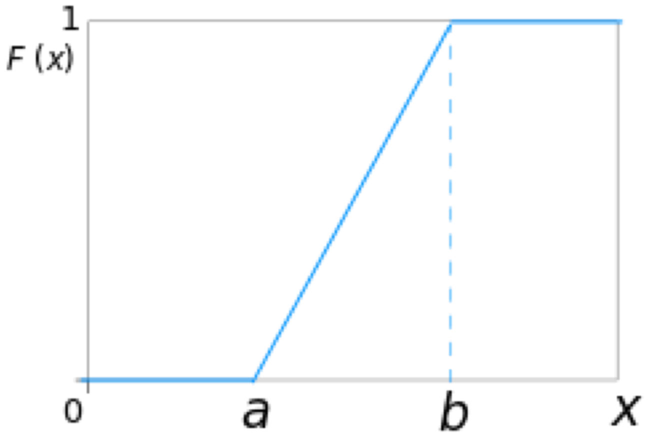
# Uniform Distribution – Discrete & Continuous

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$




$$\mathrm{E}[X] = \sum_{i=1}^{k} x_i\, p_i = x_1p_1 + x_2p_2 + \cdots + x_kp_k. \qquad \mathrm{E}[X] = \int_{\mathbb{R}} x f(x)\, dx.$$

$$\mathrm{Var}(X) = \sum_{i=1}^{n} p_i \cdot (x_i - \mu)^2, \quad \mathrm{Var}(X) = \sigma^2 = \int (x-\mu)^2 f(x)\, dx$$




| | |
|---|---|
| **Mean** | $\frac{1}{2}(a+b)$ |
| **Variance** | $\frac{1}{12}(b-a)^2$ |

# Bernoulli and Binomial Distribution

- Coin Flipping Experiment

$$\Pr(X=1) = p = 1 - \Pr(X=0) = 1 - q.$$

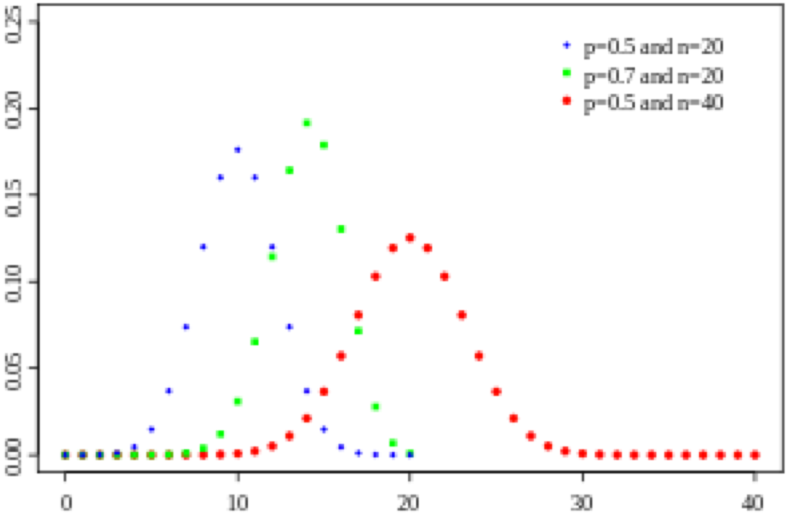
| | |
|---|---|
| **Mean** | $p$ |
| **Variance** | $p(1-p) = pq$ |

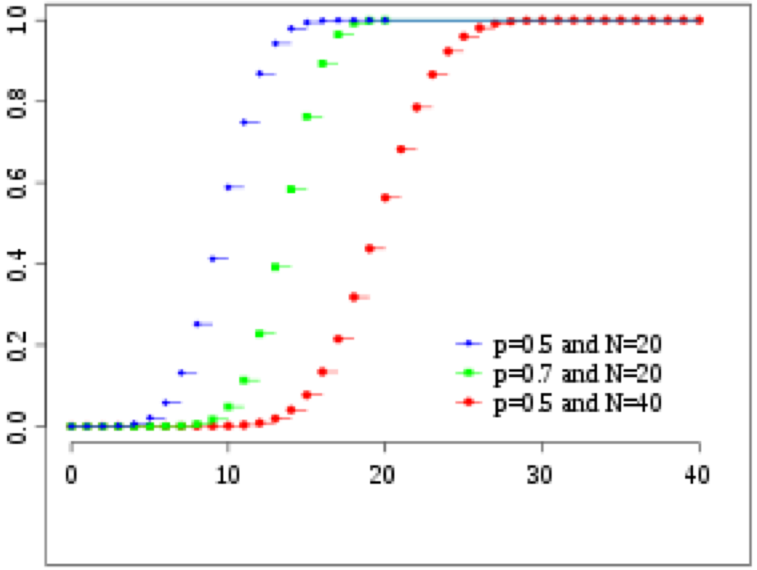
- Repeat Coin Flipping Experiment *n*-Times

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

| | |
|---|---|
| **Mean** | $np$ |
| **Variance** | $np(1-p)$ |



# Normal or Gaussian Distribution

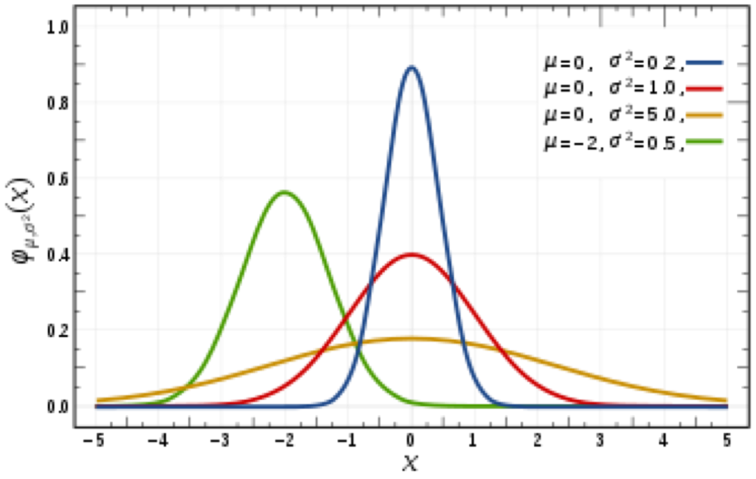
$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

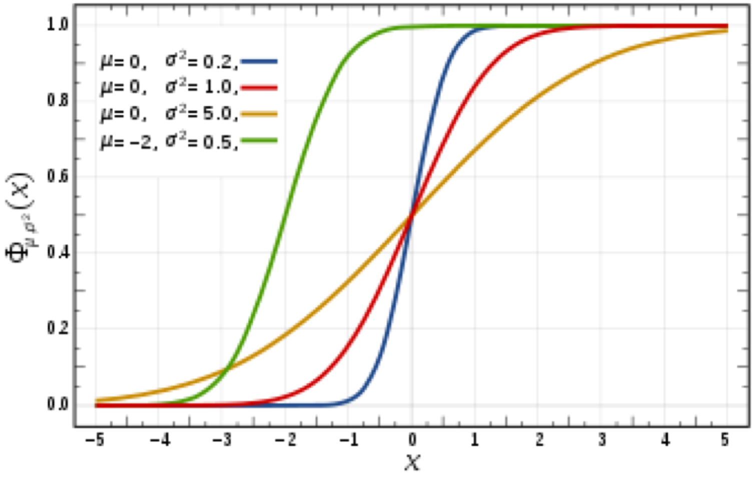
The Galton Board

## Central Limit Theorem

*by Abraham de Moivre*

When **independent random** variables are added, their properly normalized **sum** tends toward a normal distribution (informally a "bell curve") even if the original variables themselves are not normally distributed.



# Least Square

- Least Square – Linear Equation

$$\|Ax - b\|^2 = \|r\|^2 = r_1^2 + \cdots + r_m^2,$$

  - Calculus with Gradient Descent
  - Matrix Inversion with QR decomposition

$$A = QR, \quad Q^\top Q = QQ^\top = I$$

- Frequentist Treatment
  - Error Gaussian Distribution – MLE
- Bayesian Treatment
  - Expectation of Posteriors – MAP
  - Constrained Optimisation – Ridge (Gaussian), LASSO (Laplace)