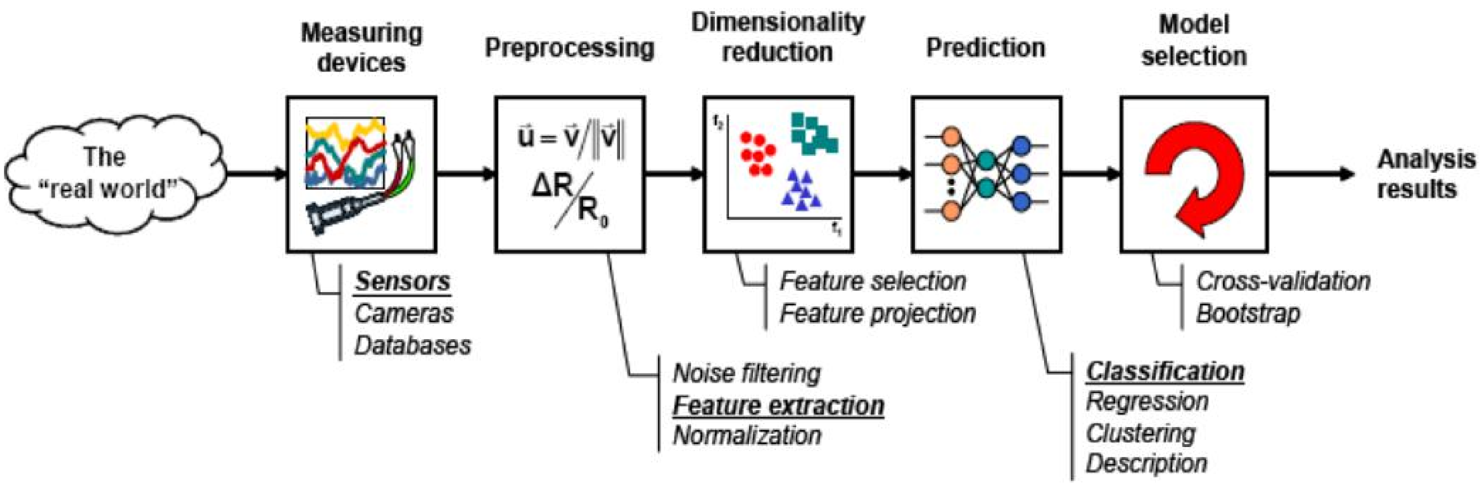
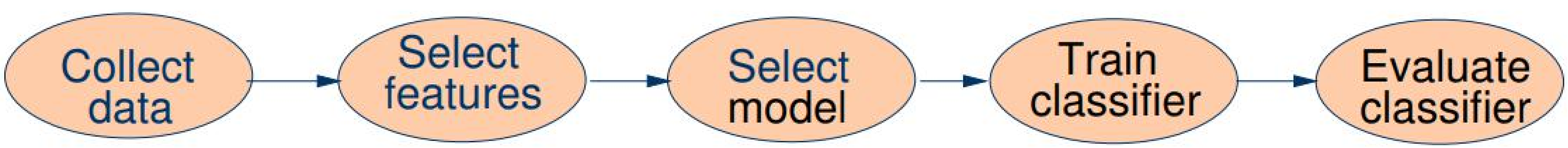
Rensselaer

Lecture 2: Probability Theory and Linear Algebra Review

Dr. Chengjiang Long
Computer Vision Researcher at Kitware Inc.
Adjunct Professor at RPI.
Email: **longc3@rpi.edu**

Recap Previous Lecture





Outline

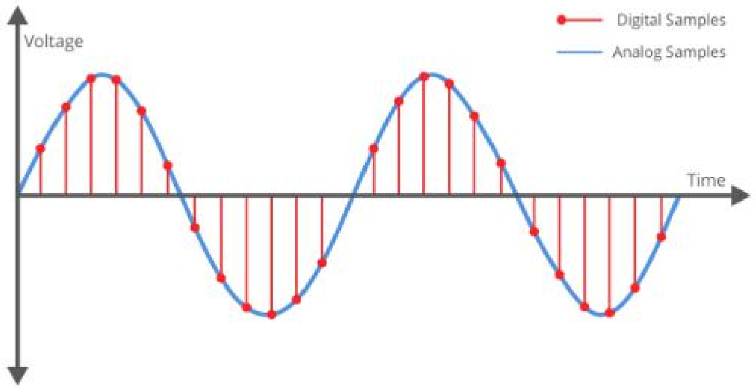
- Probability Theory Review
- Linear Algebra Review

Outline

- **Probability Theory Review**
- Linear Algebra Review

Discrete Random Variables

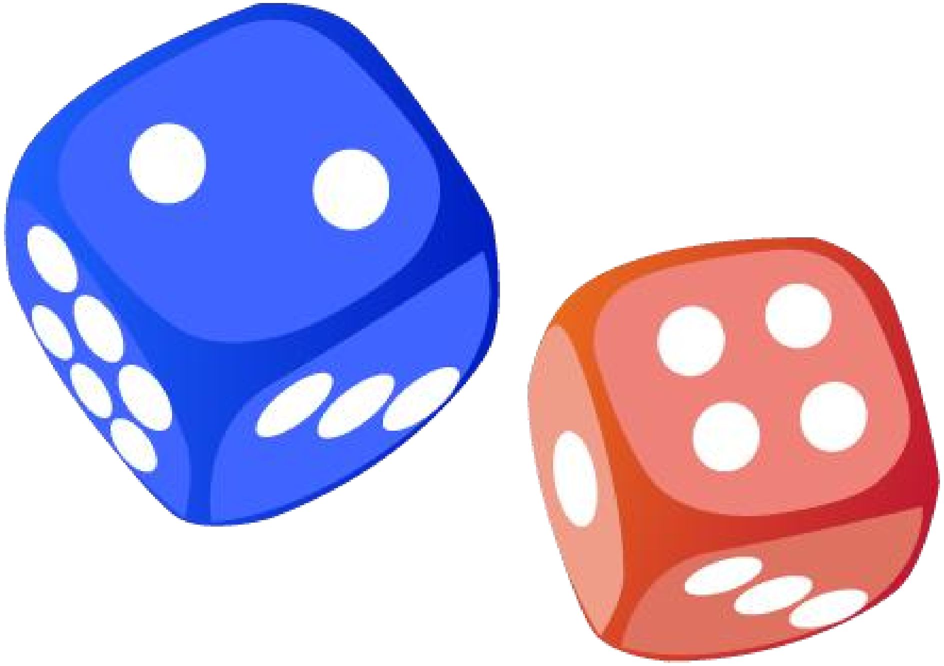
- A Random Variable is a measurement on an outcome of a random experiment.
- Discrete versus Continuous random variable: a random variable x is discrete if it can assume a finite or countably infinite number of values. x is continuous if it can assume all values in an interval.



Example

- Which of the following random variables are discrete and which are continuous？
  - x = Number of houses sold by real estate developer per week?
  - x = Number of heads in ten tosses of a coin?
  - x = Weight of a child at birth?
  - x = Time required to run100 yards?

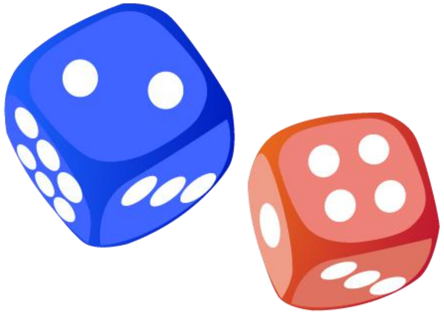
Examples



X is the Sum of Two Dice. What is the probability of X?

Probability Distribution Example: X is the Sum of Two Dice

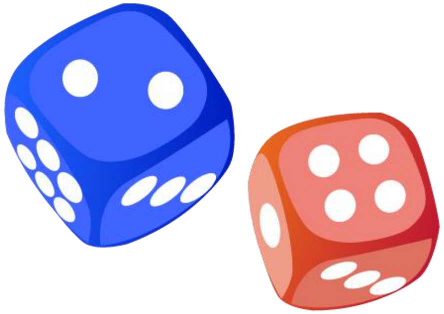
red / green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						



This sequence provides an example of a discrete random variable. Suppose that you have a red die which, when thrown, takes the numbers from 1 to 6 with equal probability.

Probability Distribution Example: X is the Sum of Two Dice

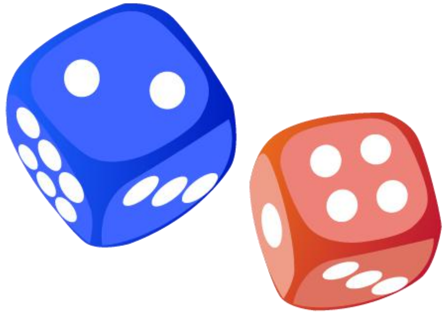
red / green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						



Suppose that you also have a green die that can take the numbers from 1 to 6 with equal probability.

Probability Distribution Example: X is the Sum of Two Dice

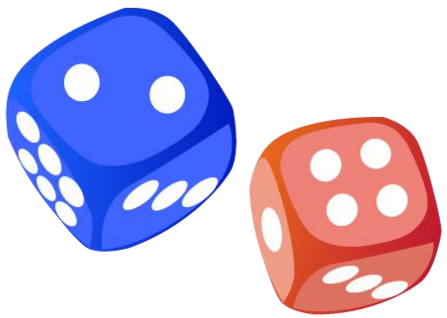
red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						



We will define a random variable X as the sum of the numbers when the dice are thrown.

Probability Distribution Example: X is the Sum of Two Dice

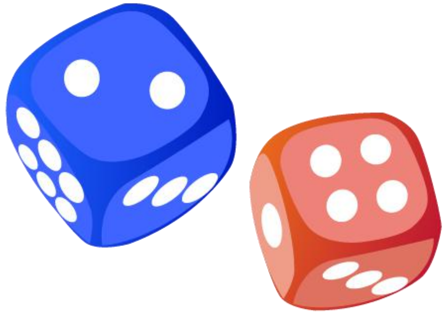
red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6				10		



For example, if the red die is 4 and the green one is 6, X is equal to 10.

Probability Distribution Example: X is the Sum of Two Dice

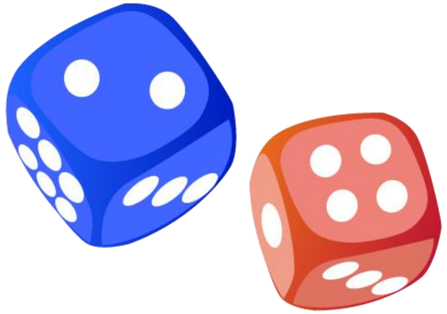
red / green	1	2	3	4	5	6
1						
2						
3						
4						
5		7				
6						



Similarly, if the red die is 2 and the green one is 5, X is equal to 7.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12



The table shows all the possible outcomes.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

We will now define f, the frequencies associated with the possible values of X.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	$\boxed{5}$	6	7
2	3	4	$\boxed{5}$	6	7	8
3	4	$\boxed{5}$	6	7	8	9
4	$\boxed{5}$	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	**4**
6	
7	
8	
9	
10	
11	
12	

For example, there are four outcomes which make X equal to 5.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Similarly you can work out the frequencies for all the other values of X.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

Finally we will derive the probability of obtaining each value of X.

Probability Distribution Example: X is the Sum of Two Dice

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

If there is 1/6 probability of obtaining each number on the red die, and the same on the green die, each outcome in the table will occur with 1/36 probability.

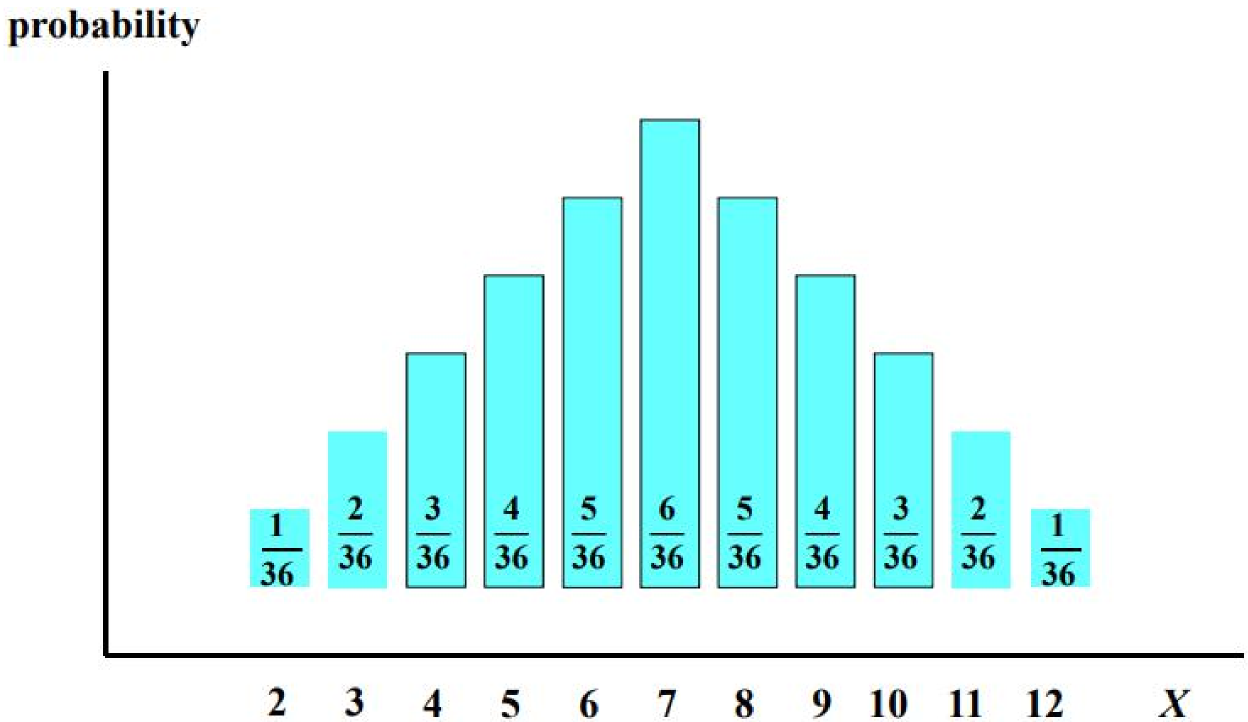
Probability Distribution Example: X is the Sum of Two Dice

green \ red	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

Hence to obtain the probabilities associated with the different values of X, we divide the frequencies by 36.

Probability Distribution Example: X is the Sum of Two Dice



The distribution is shown graphically. in this example it is symmetrical, highest for X equal to 7 and declining on either side.

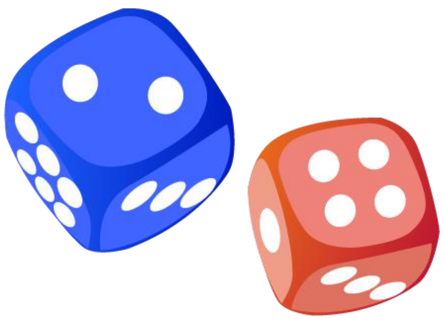
Expected Value

- Definition of E(X), the expected value of X:

$$E(X) = x_1 p_1 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

- The expected value of a random variable, also known as its population mean, is the weighted average of its possible values, the weights being the probabilities attached to the values

Expected Value Example



x_i	p_i	$x_i p_i$
x_1	p_1	$x_1 p_1$
x_2	p_2	$x_2 p_2$
x_3	p_3	$x_3 p_3$
x_4	p_4	$x_4 p_4$
x_5	p_5	$x_5 p_5$
x_6	p_6	$x_6 p_6$
x_7	p_7	$x_7 p_7$
x_8	p_8	$x_8 p_8$
x_9	p_9	$x_9 p_9$
x_{10}	p_{10}	$x_{10} p_{10}$
x_{11}	p_{11}	$x_{11} p_{11}$

$$\Sigma\, x_i p_i = E(X)$$

x_i	p_i	$x_i p_i$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36

252/36 = 7

Expected Value Properties

- Linear

$$E(X + Y) = E(X) + E(Y)$$
$$E(bX) = bE(X)$$
$$E(b) = b$$

$$\begin{aligned} Y &= b_1 + b_2X \\ E(Y) &= E(b_1 + b_2X) \\ &= E(b_1) + E(b_2X) \\ &= b_1 + b_2 E(X) \end{aligned}$$

- Also denoted by μ

Variance

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum (x_i - \mu)^2 P(X = x_i)$$

$$\mathrm{Var}(X) = \sigma^2$$

$$\mathrm{Var}(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

Pairs of Discrete Random Variables

- Let x and y be two discrete r.v.
- For each possible pair of values, we can define a joint probability ***P(x, y)***
- We can also define a joint probability mass function ***P(x,y)*** which offers a complete characterization of the pair of

$$P_x(x) = \sum_{y \in Y} P(x, y)$$

$$P_y(y) = \sum_{x \in X} P(x, y)$$

Marginal distributions

Statistical Independence

- Two random variables x and y are said to be independent, if and only if

$$P(x,y)=P_x(x)\,P_y(y)$$

that is, when knowing the value of x does not give us additional information for the value of y.

- Or, equivalently

$$E[f(x)g(y)] = E[f(x)]\,E[g(y)]$$

for any functions $f(x)$ and $g(y)$.

Conditional Probability

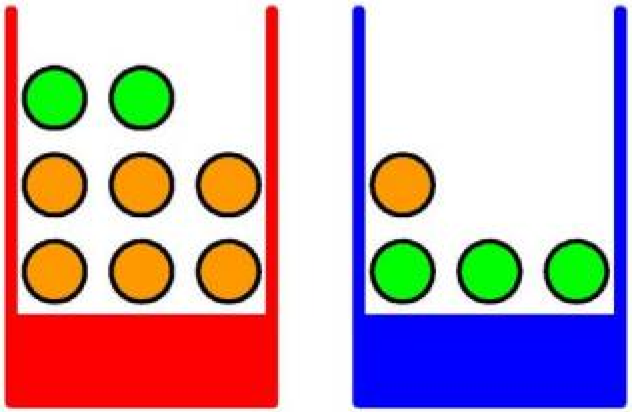
- When two r.v. are not independent, knowing one allows better estimate of the other (e.g. outside temperature, season)

$$\Pr[x = x_i \mid y = y_j] = \frac{\Pr[x = x_i , y = y_j]}{\Pr[y = y_j]}$$

- If independent $P(x|y)=P(x)$

Sum and Product Rules

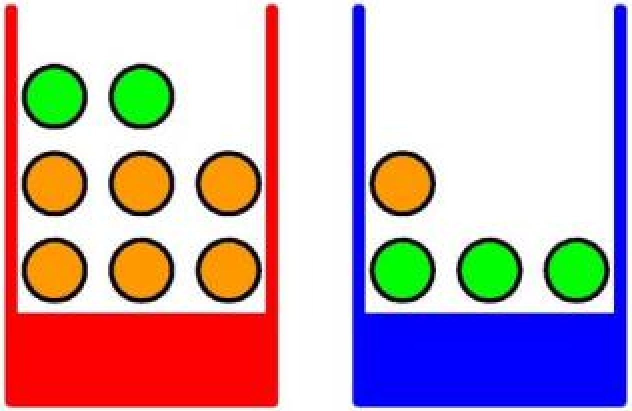
- Example:
  - We have two boxes: one red and one blue
  - Red box: 2 apples and 6 oranges
  - Blue box: 3 apples and 1 orange



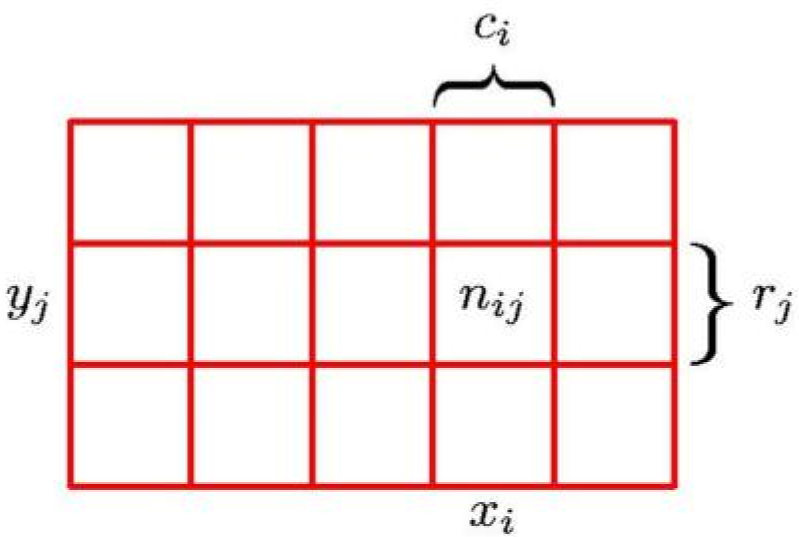
[*C.M. Bishop, "Pattern Recognition and Machine Learning", 2006*]

Sum and Product Rules

- Define:
  - B random variable for box picked (r or b)
  - F identity of fruit (a or o)
- $p(B=r)=4/10$ and $p(B=b)=6/10$
  - Events are mutually exclusive and include all possible outcomes $\Rightarrow$ their probabilities must sum to 1.



Sum and Product Rules



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

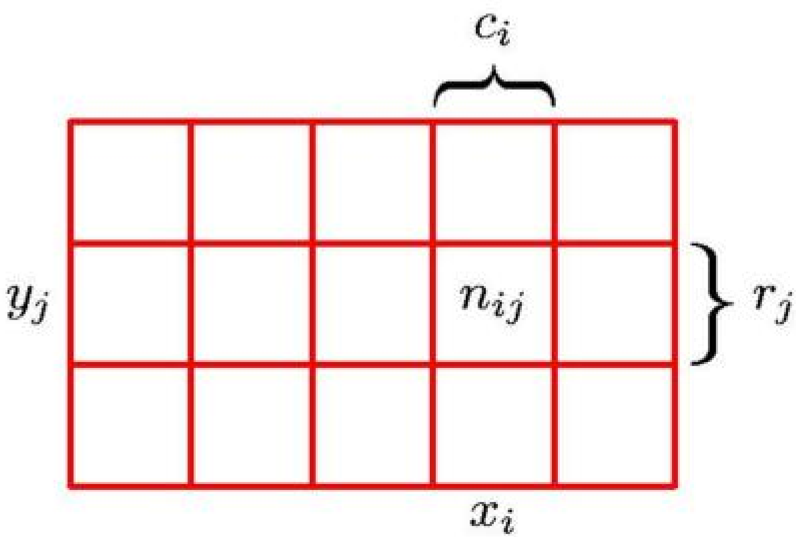
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Sum and Product Rules



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N} \sum_{j=1}^{L} n_{ij}$$

$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N}$$

$$= p(Y = y_j | X = x_i) p(X = x_i)$$

Sum and Product Rules

- Sum Rule $$p(X) = \sum_Y p(X, Y)$$
- Product Rule $$p(X, Y) = p(Y|X)p(X)$$

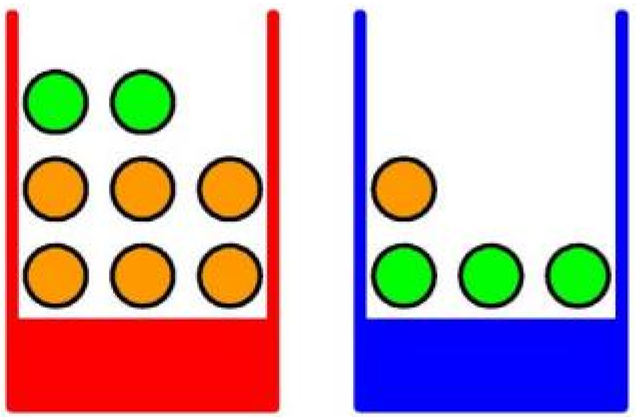
Law of Total Probability

- If an event A can occur in m different ways and if these m different ways are mutually exclusive, then the probability of A occurring is the sum of the probabilities of the sub–events

$$P(X = x_i) = \sum_j P(X = x_i \mid Y = y_j) P(Y = y_j)$$

Sum and Product Rules

- Back to the fruit baskets
  - $p(B=r)=4/10$ and $p(B=b)=6/10$
  - $p(B=r) + p(B=b) = 1$
- Conditional probabilities
  - $p(F=a \mid B = r) = 1/4$
  - $p(F=o \mid B = r) = 3/4$
  - $p(F=a \mid B = b) = 3/4$
  - $p(F=o \mid B = b) = 1/4$



Sum and Product Rules

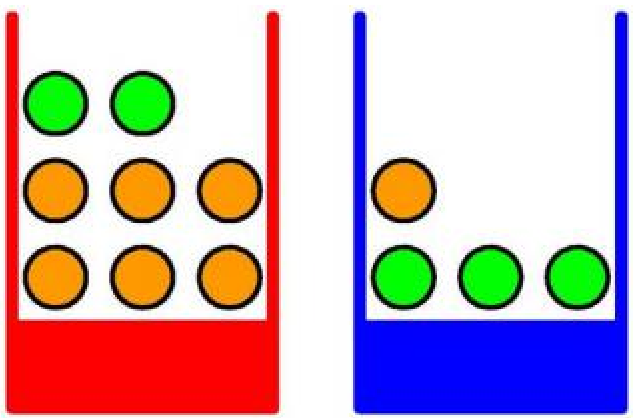
- Note:

$p(F{=}a \mid B = r) + p(F{=}o \mid B = r) = 1$

$p(F{=}a) = p(F{=}a \mid B = r)\, p(B{=}r) + p(F{=}a \mid B = b) p(B{=}b)$

$= 1/4 * 4/10 + 3/4 * 6/10 = 11/20$

- Sum rule: $p(F{=}o) = ?$



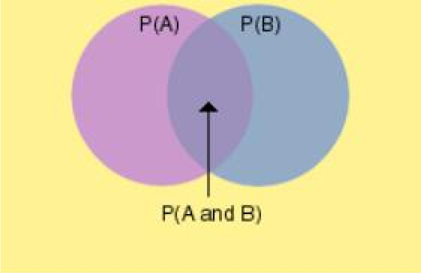
Conditional Probability Example

- A jar contains black and white marbles.
  - Two marbles are chosen without replacement.
  - The probability of selecting a black marble and then a white marble is 0.34.
  - The probability of selecting a black marble on the first draw is 0.47.

- What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

$$P(White \mid Black) = \frac{P(Black \wedge White)}{P(Black)} = \frac{0.34}{0.47} = 0.72$$

A is black in first draw, B is white in second draw



Law of Total Probability

$$P_x(x) = \sum_{y \in Y} P(x, y)$$

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

Bayes Rule

$$P(x \mid y) = \frac{P(x, y)}{P(y)} = \frac{P(y \mid x) P(x)}{\sum_{x \in X} P(x, y)}$$

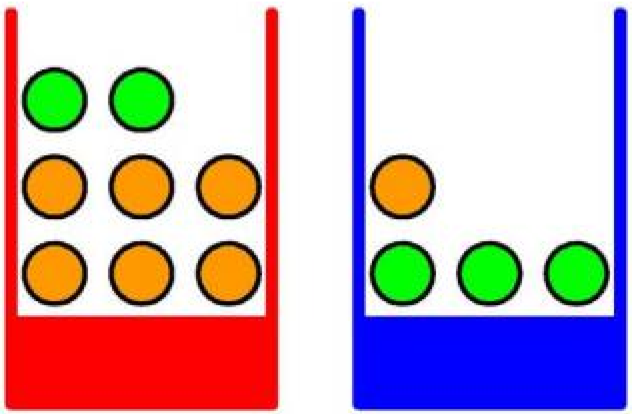
$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

- x is the unknown cause
- y is the observed evidence
- Bayes rule shows how probability of x changes after we have observed y

Bayes Rule on the Fruit Example

- Suppose we have selected an orange. Which box did it come from？

$$p(B=r\mid F=o)=\frac{p(F=o\mid B=r)p(B=r)}{p(F=o)}=\frac{\frac{3}{4}\times\frac{4}{10}}{\frac{9}{20}}=\frac{2}{3}$$



Continuous Random Variables

- Examples: room temperature, time to run 100m, weight of child at birth…
- Cannot talk about probability of that x has a particular value
- Instead, probability that x falls in an interval ⇨ **probability density function**

$$\Pr[x \in (a,b)] = \int_a^b p(x)dx$$

$$p(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} p(x)dx = 1$$

Expected Value

$$E[x] = \mu = \int_{-\infty}^{\infty} xp(x)dx$$

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

$$Var[x] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x)dx$$

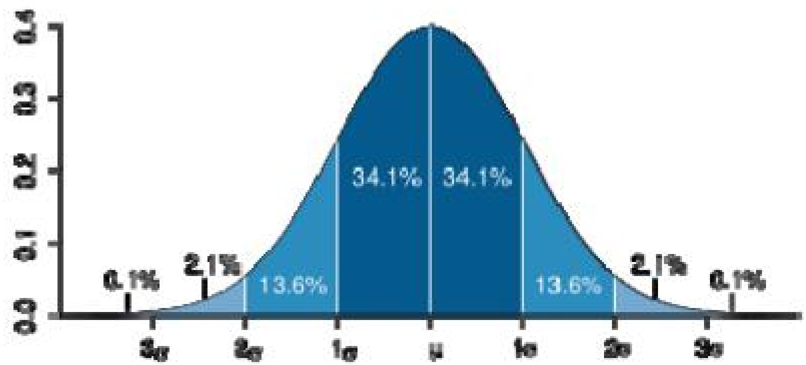
- Bayes rule $p(x \mid y) = \dfrac{p(y \mid x)p(x)}{\int_{-\infty}^{\infty} p(y \mid x)p(x)dx}$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

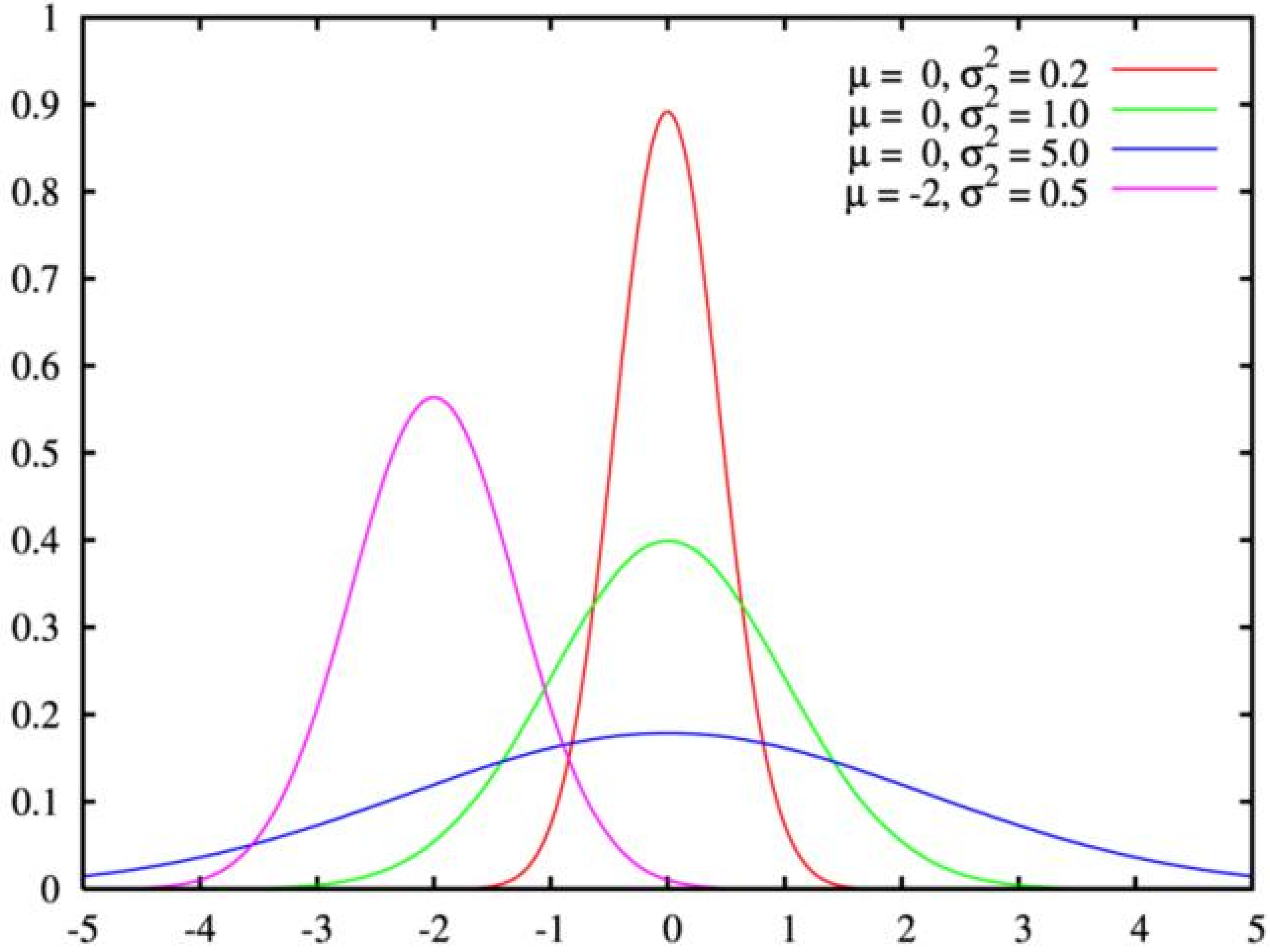
Normal (Gaussian) Distribution

- Central Limit Theorem: under various conditions, the distribution of the sum of d independent random variables approaches a limiting form known as the **normal distribution**

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

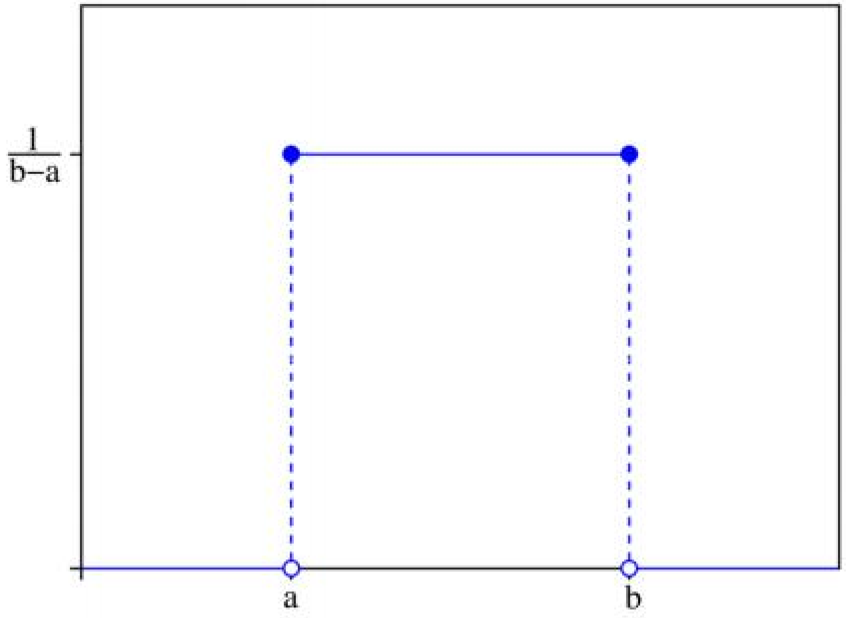


Normal (Gaussian) Distribution



Uniform Distribution

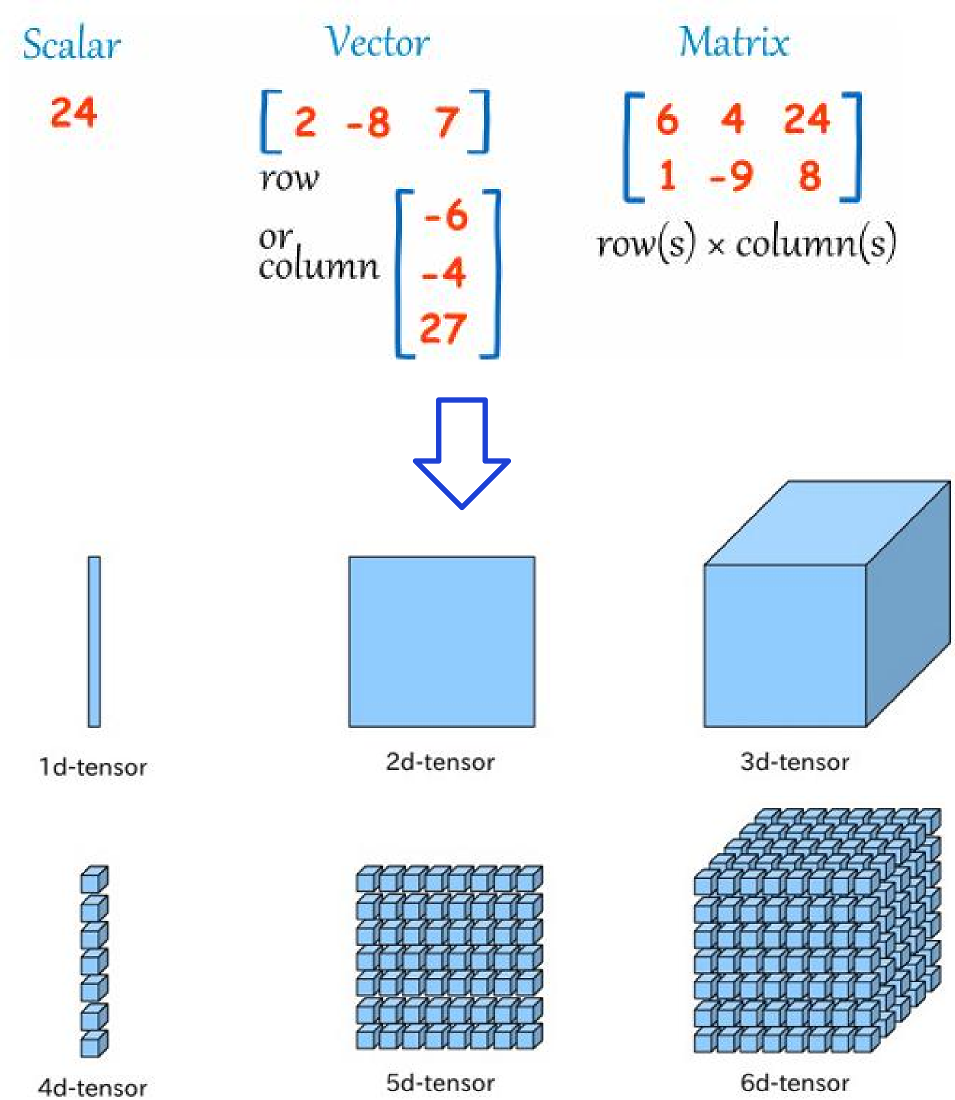
$$p(x)=\begin{cases} \dfrac{1}{b-a} & \text{for } a<=x<=b \\ 0 & \text{for } x<a \text{ or } x>b \end{cases}$$



Outline

- Probability Theory Review
- Linear Algebra Review
- Summary

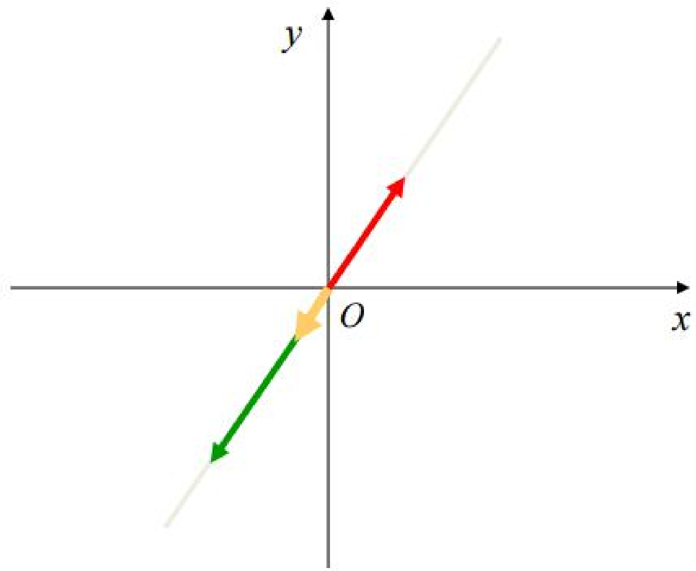
Linear Algebra



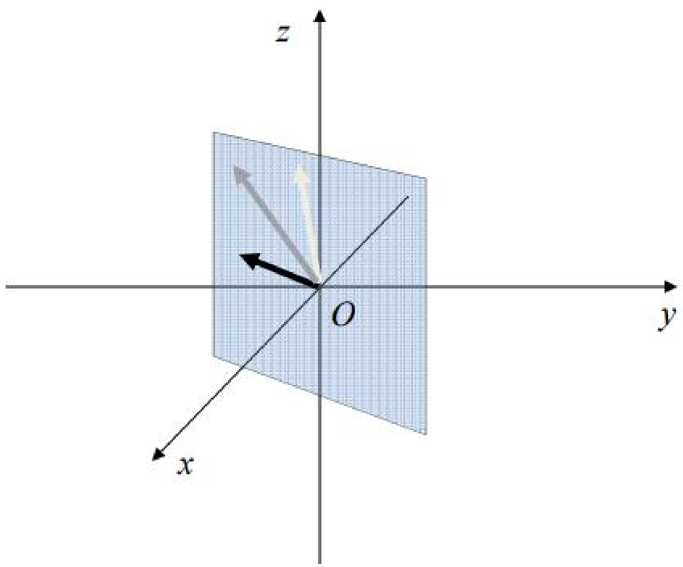
Vector space

- Informal definition:
  - $V \neq \emptyset$ (a non-empty set of vectors)
  - $\mathbf{v}, \mathbf{w} \in V \Rightarrow \mathbf{v} + \mathbf{w} \in V$ (closed under addition)
  - $\mathbf{v} \in V$, α is scalar $\Rightarrow \alpha\mathbf{v} \in V$ (closed under multiplication by scalar)
- Formal definition includes axioms about associativity and distributivity of the $+$ and $\bullet$ operators.
- $0 \in V$ Always!!

Example: Linear subspace of R^2 and R^3



Line



Plane

Linear independence

- The vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ are a linearly independent set if:

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_k\mathbf{v}_k = 0 \quad \Leftrightarrow \quad \alpha_i = 0 \;\forall i$$

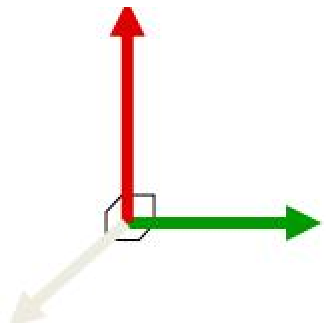
- **It means that none of the vectors can be obtained as a linear combination of the others.**

Example

- Parallel vectors are always dependent:

$$\mathbf{v} = 2.4\,\mathbf{w} \;\Rightarrow\; \mathbf{v} + (-2.4)\mathbf{w} = 0$$

- Orthogonal vectors are always linearly independent:



Basis of Vector

- $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ are **linearly independent**.

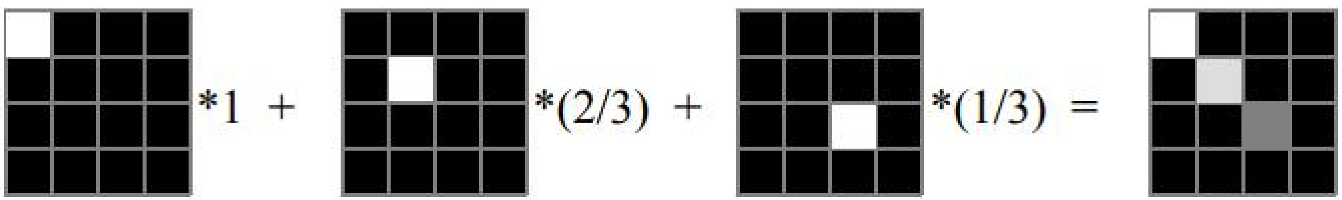
- $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ **span** the whole vector space ***V***:

$$V = \{\alpha_I \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n \mid \alpha_I \text{ scalars}\}$$

- Any vector in ***V*** is a **unique** linear combination of the basis.

- The number of basis vectors is called the **dimension** of ***V***.

Example



Note that a matrix can be interpreted as a vector.

Matrix representation

- Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis of V.
- Every $\mathbf{v} \in V$ has a unique representation.

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \ldots + \alpha_n \mathbf{v}_n$$

- Denote v by the column-vector:

$$\mathbf{v} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

- The basis vectors are therefore denoted:

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

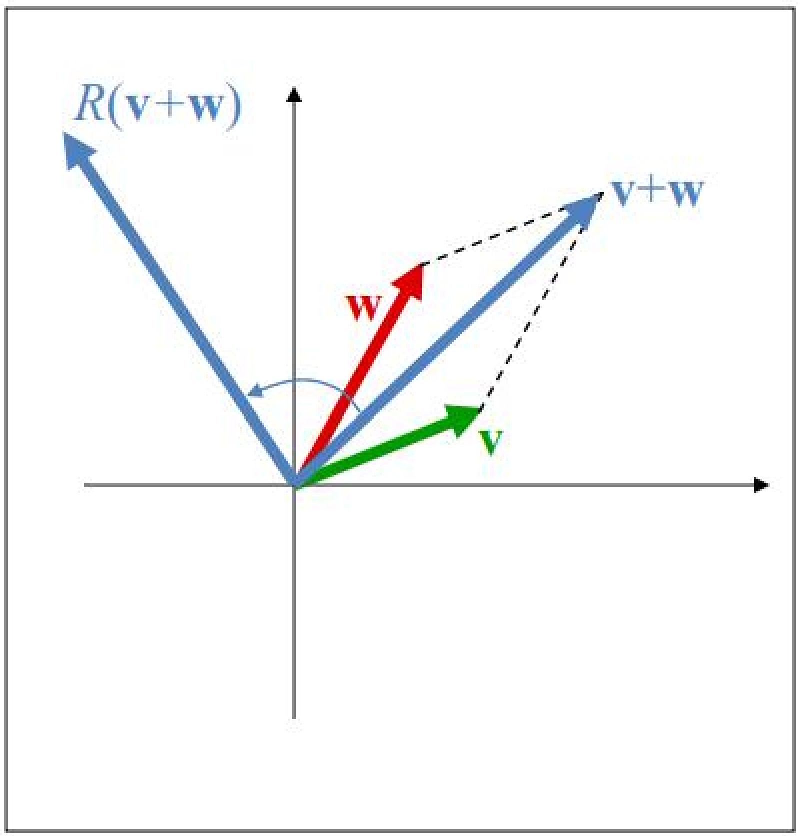
Linear Operation

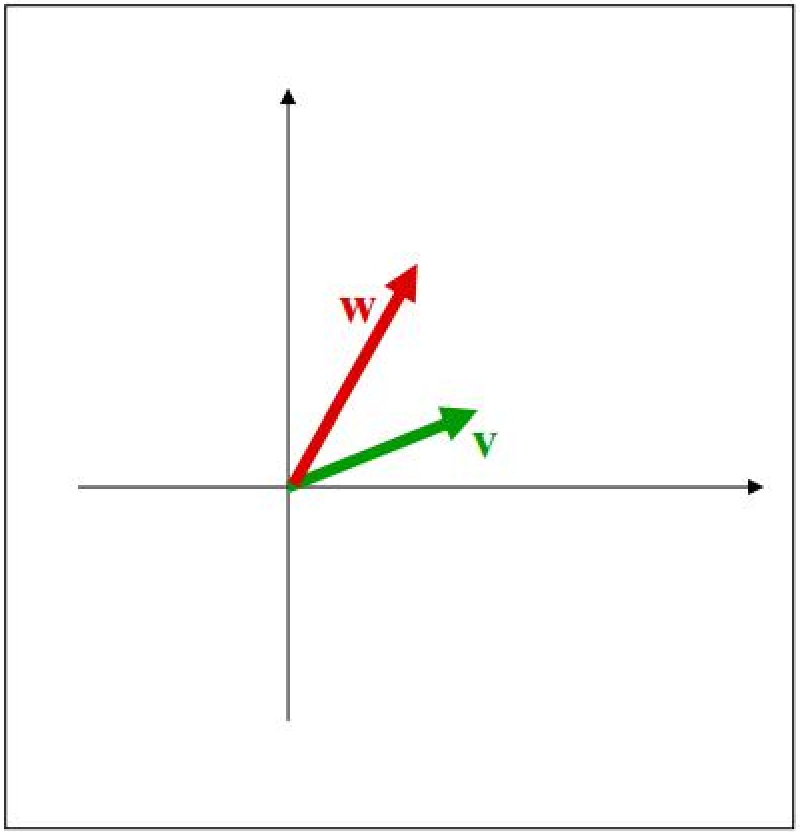
- $A : V \to W$ is called linear operator if:

$$A(\mathbf{v} + \mathbf{w}) = A(\mathbf{v}) + A(\mathbf{w})$$
$$A(\alpha\,\mathbf{v}) = \alpha\, A(\mathbf{v})$$

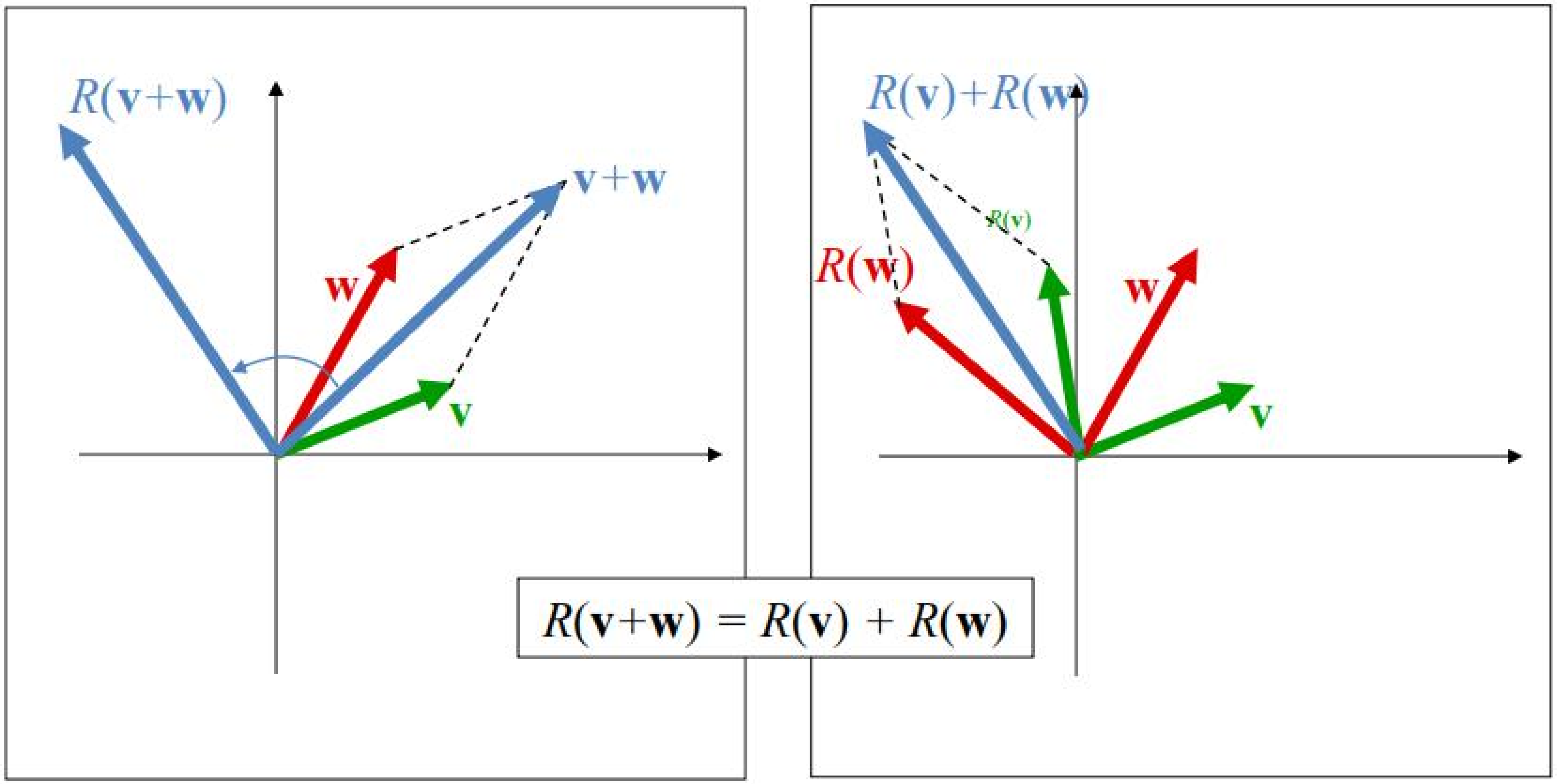
- In particular, $A(\mathbf{0}) = \mathbf{0}$
- Linear operators we know:
  – Scaling
  – Rotation, reflection
  – **Translation is not linear – moves the origin**

Example: Rotation is a linear operator





Example: Rotation is a linear operator



Matrix operations

- Addition, subtraction, scalar multiplication – simple…
- Multiplication of matrix by column vector:

$$\underset{A}{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}} \underset{\mathbf{b}}{\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}} = \begin{pmatrix} \sum_i a_{1i} b_i \\ \vdots \\ \sum_i a_{mi} b_i \end{pmatrix}$$

Matrix by vector multiplication

- Sometimes a better way to look at it:
  - *Ab* is a linear combination of *A*'s ***columns***

$$\begin{pmatrix} | & | & & | \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = b_1 \begin{pmatrix} | \\ \mathbf{a}_1 \\ | \end{pmatrix} + b_2 \begin{pmatrix} | \\ \mathbf{a}_2 \\ | \end{pmatrix} + \ldots + b_n \begin{pmatrix} | \\ \mathbf{a}_n \\ | \end{pmatrix}$$

How to Understand Matrix Product

$$\begin{cases} a_{11}x_1 + a_{12}x_2 = y_1 \\ a_{21}x_1 + a_{22}x_2 = y_2 \end{cases} \qquad \begin{cases} b_{11}t_1 + b_{12}t_2 = x_1 \\ b_{21}t_1 + b_{22}t_2 = x_2 \end{cases}$$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \qquad \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\Downarrow$$

$$\boxed{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

$$\Downarrow$$

$$\begin{cases} a_{11}(b_{11}t_1 + b_{12}t_2) + a_{12}(b_{21}t_1 + b_{22}t_2) = y_1 \\ a_{21}(b_{11}t_1 + b_{12}t_2) + a_{22}(b_{21}t_1 + b_{22}t_2) = y_2 \end{cases}$$

$$\Downarrow$$

$$\begin{cases} (a_{11}b_{11} + a_{12}b_{21})t_1 + (a_{11}b_{12} + a_{12}b_{22})t_2 = y_1 \\ (a_{21}b_{11} + a_{22}b_{21})t_1 + (a_{21}b_{12} + a_{22}b_{22})t_2 = y_2 \end{cases}$$

$$\boxed{\begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}} \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

Transposition: make the rows to be the columns

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & \cdots & a_{m1} \\ \vdots & \ddots & \vdots \\ a_{1n} & \cdots & a_{mn} \end{pmatrix}$$

$$(AB)^T = B^T A^T$$

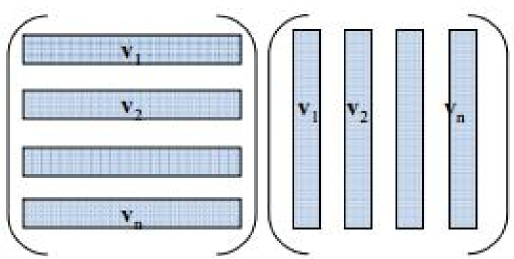
Matrix properties

- Matrix $A_{(n \times n)}$ is non–singular if $\exists B, AB = BA = I$
- $B = A^{-1}$ is called the inverse of ***A***.
- A is non–singular $\Leftrightarrow$ $det\ A \neq 0$
- If A is non–singular then the equation $A\mathbf{x} = \mathbf{b}$ has one unique solution for each b.
- A is non–singular $\Leftrightarrow$ the rows of A are linearly independent (and so are the columns)

Orthogonal matrices

- Matrix $A_{(n \times n)}$ is orthogonal if $A^{-1} = A^{T}$.
- Follows: $AA^{T} = A^{T}A = I$
- The rows of A are orthonormal vectors!
- Proof:

$$I = A^{T}A =$$



$$= \begin{pmatrix} \mathbf{v}_{\mathrm{i}}{}^{T}\mathbf{v}_{\mathrm{j}} \end{pmatrix} = \begin{pmatrix} \delta_{ij} \end{pmatrix}$$

$$\Rightarrow \langle \mathbf{v}_{\mathrm{i}}, \mathbf{v}_{\mathrm{i}} \rangle = 1 \quad \Rightarrow \quad \|\mathbf{v}_{\mathrm{i}}\| = 1; \quad \langle \mathbf{v}_{\mathrm{i}}, \mathbf{v}_{\mathrm{j}} \rangle = 0$$

The Trace

- The trace of a square matrix denoted by tr(A) is the sum of the diagonal elements

$$\operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}$$

- Some properties

$$
\begin{aligned}
\operatorname{Tr}(\mathbf{A}+\mathbf{B}) &= \operatorname{Tr}(\mathbf{A}) + \operatorname{Tr}(\mathbf{B}) \\
\operatorname{Tr}(\mathbf{E}) &= n \qquad \text{(trace of identity matrix)} \\
\operatorname{Tr}(\mathbf{O}) &= 0 \qquad \text{(trace of zero matrix)} \\
\operatorname{Tr}(\mathbf{ABC}) &= \operatorname{Tr}(\mathbf{CAB}) = \operatorname{Tr}(\mathbf{BCA}) \\
\operatorname{Tr}(c\mathbf{A}) &= c\operatorname{Tr}(\mathbf{A}) \quad c \in \mathbb{C} \\
\operatorname{Tr}(\mathbf{A}^{\mathrm{T}}) &= \operatorname{Tr}(\mathbf{A})
\end{aligned}
$$

Example

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 3 \\ 11 & 5 & 2 \\ 6 & 12 & -5 \end{pmatrix}$$

$$\operatorname{tr}(A) = \sum_{i=1}^{3} a_{ii} = a_{11} + a_{22} + a_{33} = -1 + 5 + (-5) = -1.$$

The Determinant

(1.) A multiple of one row of "A" is added to another row to produce a matrix, "B", then: $|\mathbf{A}| = |\mathbf{B}|$.

(2.) If two rows are interchanged to produce a matrix, "B", then: $|\mathbf{B}| = -|\mathbf{A}|$.

(3.) If one row is multiplied by "k" to produce a matrix, "B", then: $|\mathbf{B}| = \mathbf{k}\cdot|\mathbf{A}|$.

(4.) If "A" and "B" are both n x n matrices then: $|\mathbf{A}\cdot\mathbf{B}| = |\mathbf{A}|\cdot|\mathbf{B}|$.

(5.) $\left|\mathbf{A}^{\mathbf{T}}\right| = |\mathbf{A}|$.

$$|\mathbf{A}| = \left|\begin{pmatrix} 1 & 5 & -6 \\ -1 & -4 & 4 \\ -2 & -7 & 9 \end{pmatrix}\right| = \left|\begin{pmatrix} 1 & 5 & -6 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix}\right| = \left|\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}\right| = (1)\cdot(1)\cdot(3)\cdot|\mathbf{I}| = 3$$

The Determinant

- For a square matrix A, the determinant is denoted by |A| or det(A)

$$
\begin{aligned}
|A| &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A_{\backslash i, \backslash j}| \qquad (\text{for any } j \in 1, \ldots, n) \\
&= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A_{\backslash i, \backslash j}| \qquad (\text{for any } i \in 1, \ldots, n)
\end{aligned}
$$

$$
\begin{aligned}
|[a_{11}]| &= a_{11} \\
\left| \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \right| &= a_{11}a_{22} - a_{12}a_{21} \\
\left| \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \right| &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\
&\quad - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}
\end{aligned}
$$

$|A| = |A^T|$

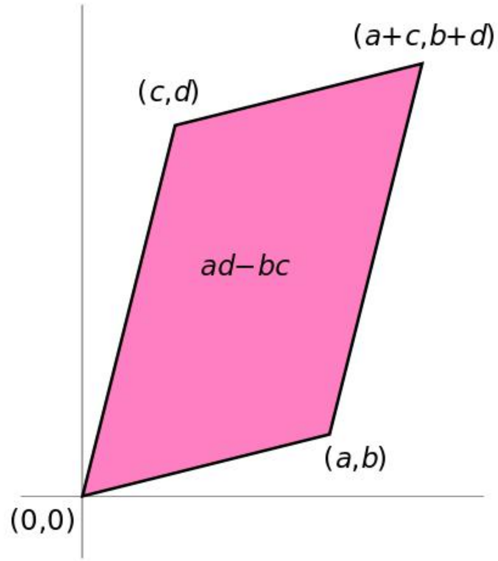
$|AB| = |A|\,|B|$

$|A| = 0$, if and only if A is singular

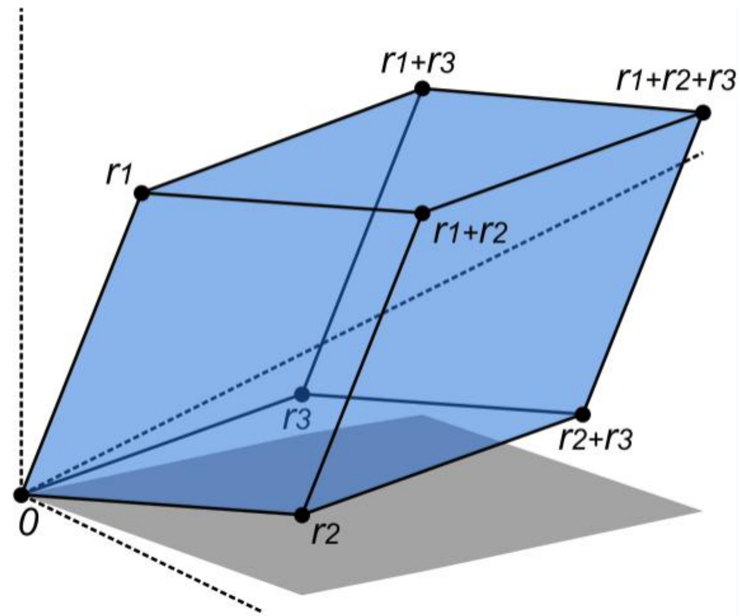
– Else, $|A^{-1}| = 1/|A|$

The Determinant

For proof, see https://math.stackexchange.com/questions/427528/why-determinant-is-volume-of-parallelepiped-in-any-dimensions



The area of the parallelogram is the absolute value of the determinant of the matrix formed by the vectors representing the parallelogram's sides.



The volume of this parallelepiped is the absolute value of the determinant of the matrix formed by the rows constructed from the vectors r1, r2, and r3.

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

$$\begin{aligned} \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} &= a\begin{vmatrix} e & f \\ h & i \end{vmatrix} - b\begin{vmatrix} d & f \\ g & i \end{vmatrix} + c\begin{vmatrix} d & e \\ g & h \end{vmatrix} \\ &= a(ei - fh) - b(di - fg) + c(dh - eg) \\ &= aei + bfg + cdh - ceg - bdi - afh. \end{aligned}$$

Covariance

- Covariance is a numerical measure that shows how much two random variables change together

$$\sigma_{jk} = E\left[(Y_{ij} - \mu_j)(Y_{ik} - \mu_k)\right]$$

- Positive covariance: if one increases, the other is likely to increase
- Negative covariance: …
- More precisely: **the covariance is a measure of the linear dependence between the two variables**

Covariance Matrix

- For a vector of repeated measures, we define the covariance matrix:

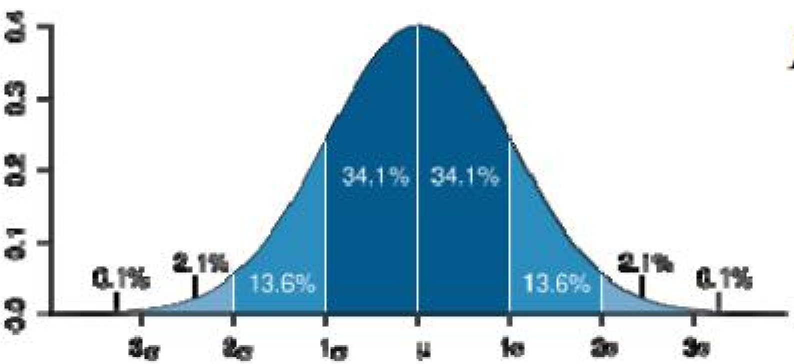
$$\mathrm{Cov}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{in} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(Y_{i1}) & \mathrm{Cov}(Y_{i1}, Y_{i2}) & \cdots & \mathrm{Cov}(Y_{i1}, Y_{in}) \\ \mathrm{Cov}(Y_{i2}, Y_{i1}) & \mathrm{Var}(Y_{i2}) & \cdots & \mathrm{Cov}(Y_{i2}, Y_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(Y_{in}, Y_{i1}) & \mathrm{Cov}(Y_{in}, Y_{i2}) & \cdots & \mathrm{Var}(Y_{in}) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix},$$

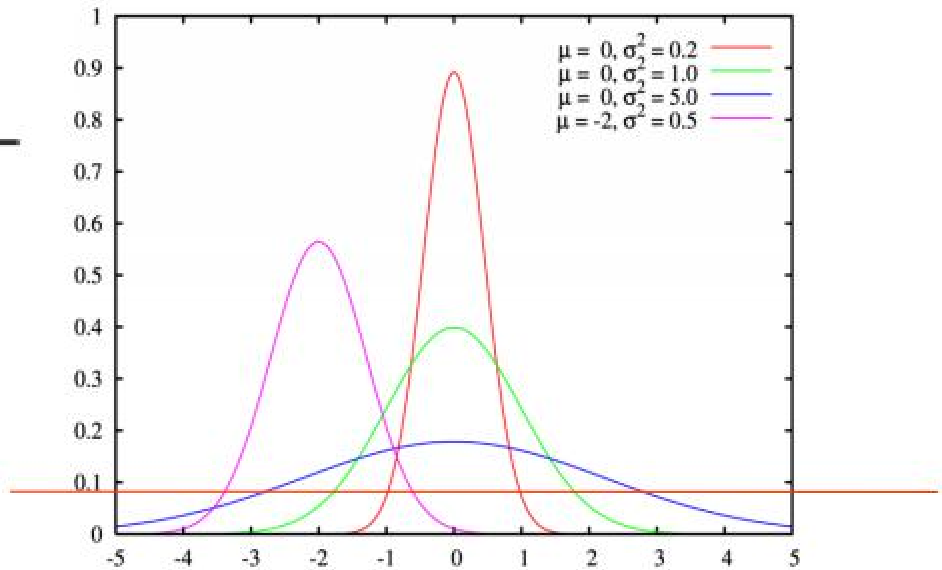
where $\mathrm{Cov}\,(Y_{ij}, Y_{ik}) = \sigma_{jk} = \sigma_{kj} = \mathrm{Cov}(Y_{ik}, Y_{ij})$.

- It is a symmetric, square matrix.

Normal Distribution



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



Multivariate Normal Distribution

- The multivariate normal density in d dimensions is:

$$P(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]$$

where:

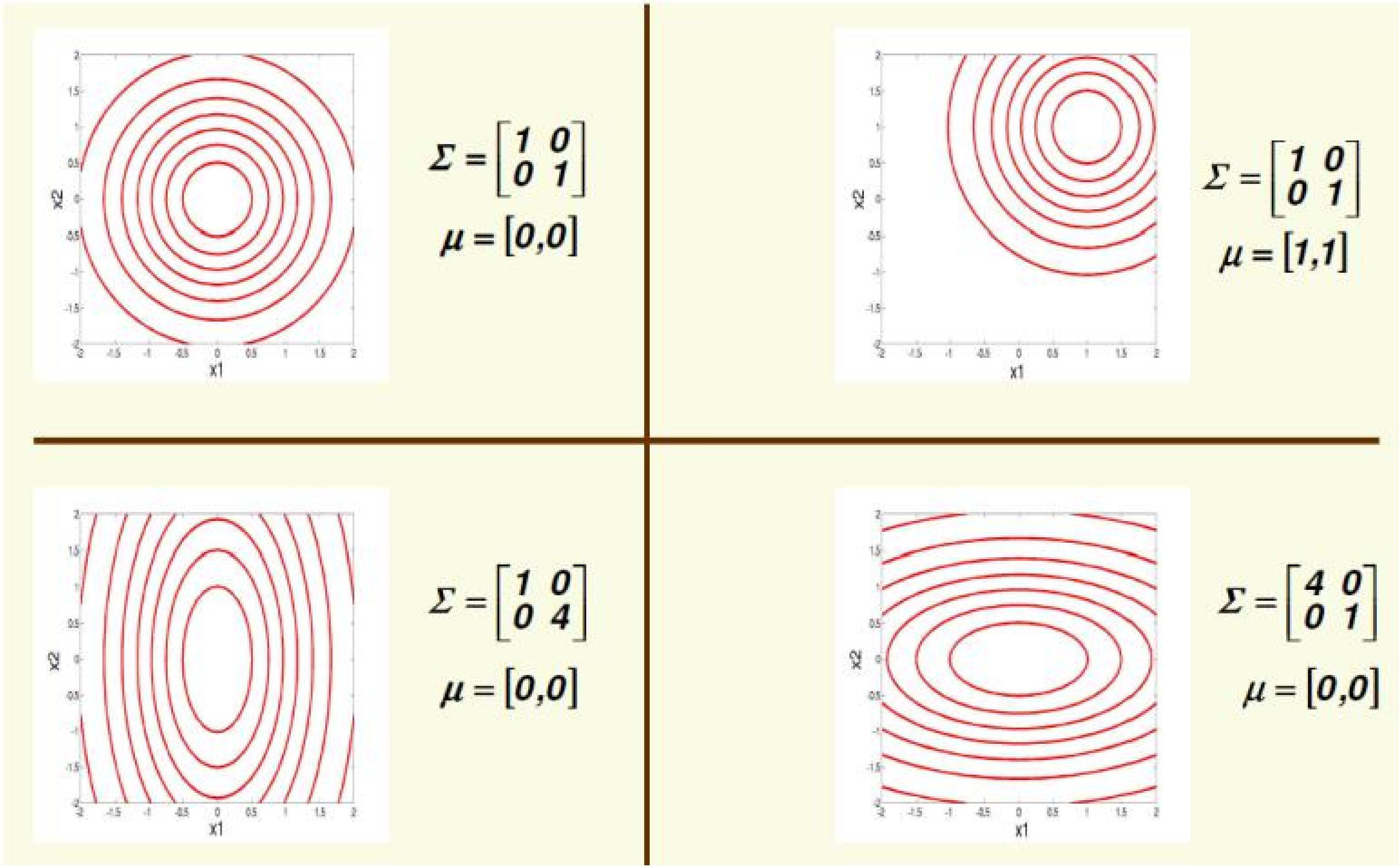
$x = (x_1, x_2, \ldots, x_d)^t$

$\mu = (\mu_1, \mu_2, \ldots, \mu_d)^t$ mean vector

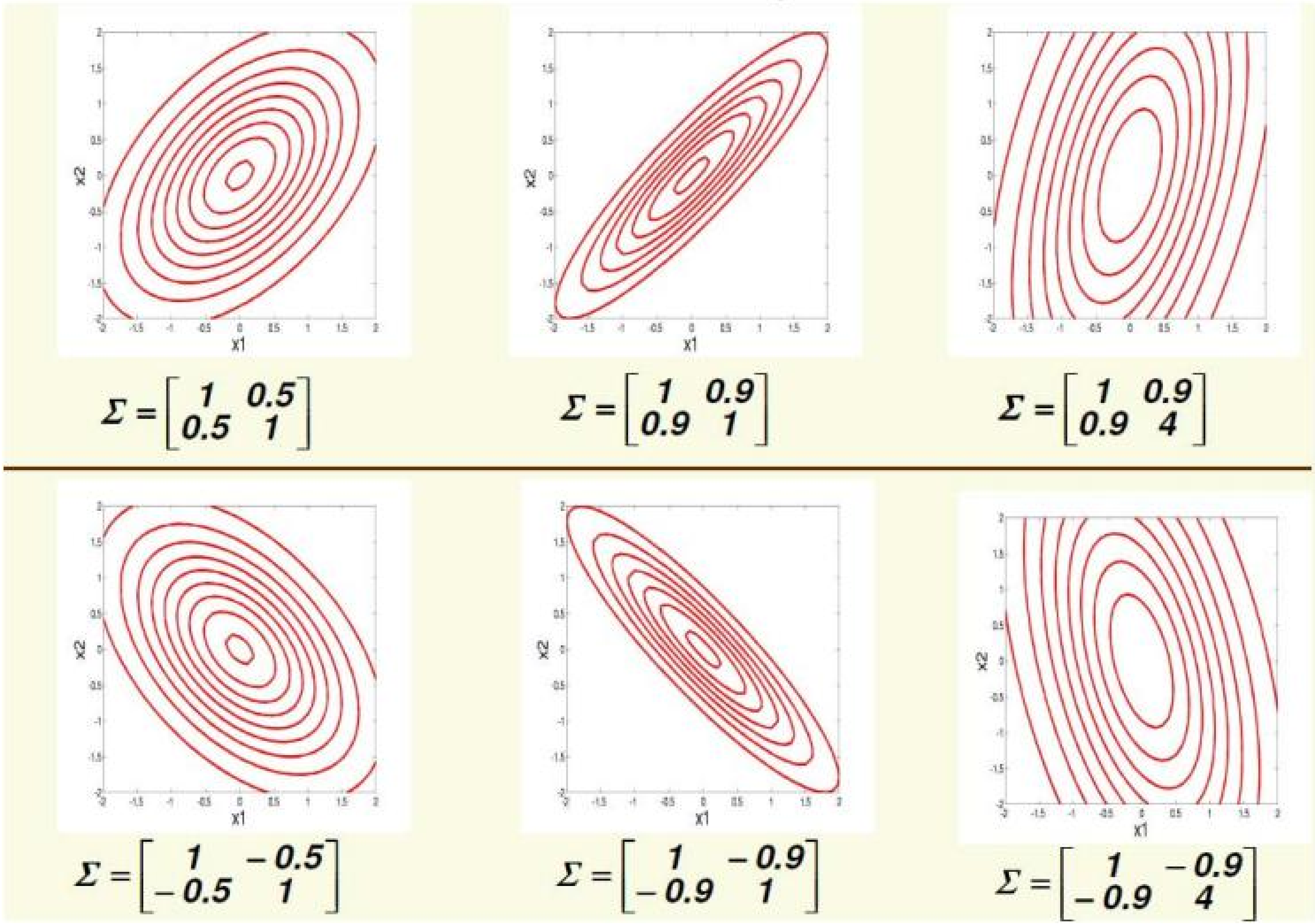
$\Sigma = d \times d$ covariance matrix

$|\Sigma|$ and Σ^{-1} are the determinant and inverse respectively

Multivariate Normal Distribution



Multivariate Normal Distribution



Eigenvalues and Eigenvectors

- For an $n \times n$ square matrix A, e is an eigenvector with eigenvalue λ if

$$Ae = \lambda e$$

Or

$$(A - \lambda I)e = 0$$

- If $(A - \lambda I)$ is invertible, the only solution is $e = 0$ (trivial)
- For non–trivial solutions: $\det(A - \lambda I) = 0$
- Above equation is called the "characteristic polynomial"
- Solutions are not unique
  – If e is an eigenvector αe is also an eigenvector

Example: a 2×2 matrix

$$\det[\mathbf{A}-\lambda\mathbf{I}]=\begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{vmatrix}=(a_{11}-\lambda)(a_{22}-\lambda)-a_{12}a_{21}=0$$

$$0=a_{11}a_{22}-a_{12}a_{21}-\lambda(a_{11}+a_{22})+\lambda^2$$

$$\mathbf{A}=\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$0=a_{11}a_{22}-a_{12}a_{21}-(a_{11}+a_{22})\lambda+\lambda^2$$

$$0=1\cdot 4-2\cdot 2-(1+4)\lambda+\lambda^2$$

$$(1+4)\lambda=\lambda^2$$

$$\Downarrow$$

$\boxed{\lambda=0}$ and $\lambda=5$

$$\mathbf{A}\mathbf{x}=\lambda\mathbf{x},\quad (\mathbf{A}-\lambda\mathbf{I})\mathbf{x}=0$$

$$\left[\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}-\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}\right]\cdot\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}\cdot\begin{bmatrix} x \\ y \end{bmatrix}=\begin{bmatrix} 1x+2y \\ 2x+4y \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

x=2, y=-1

Example: a 2×2 matrix

$$\det[\mathbf{A}-\lambda\mathbf{I}] = \begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{vmatrix} = (a_{11}-\lambda)(a_{22}-\lambda) - a_{12}a_{21} = 0$$

$$0 = a_{11}a_{22} - a_{12}a_{21} - \lambda(a_{11}+a_{22}) + \lambda^2$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

$$0 = a_{11}a_{22} - a_{12}a_{21} - (a_{11}+a_{22})\lambda + \lambda^2$$

$$0 = 1\cdot 4 - 2\cdot 2 - (1+4)\lambda + \lambda^2$$

$$(1+4)\lambda = \lambda^2$$

⇩

λ=0 and $\boxed{\lambda=5}$

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \quad (\mathbf{A}-\lambda\mathbf{I})\mathbf{x} = 0$$

$$\left[\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} - \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix}\right] \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -4 & 2 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -4x+2y \\ 2x-1y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

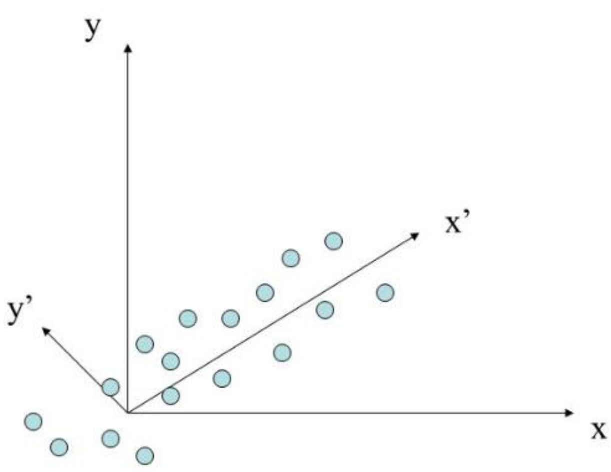
x=1, y=2

Properties

- The product of the eigenvalues = |A|
- The sum of the eigenvalues = trace(A)
- A symmetric matrix has real eigenvalues.
- A real symmetric matrix can be written as:

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{e}_i \mathbf{e}_i^\mathsf{T}$$

$$\mathbf{A} = \mathbf{E} \Lambda \mathbf{E}^\mathsf{T}$$



Q & A