COMP 411/511
Computer Vision with Deep Learning

Probability Tutorial

Agenda

- Introduction to Probability
- Conditional Probability
- Random Variables
- Probability Distributions

Adapted from *Stanford CS229 Probability Refresher* and *Deep Learning Book, Chapter 3*

Introduction to Probability

- **Sample Space *S***
  The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by ***S***.

- Examples
  - Coin toss: {H, T}
  - 2 Coin toss: $\{(H_1, H_2), (H_1, T_2), (T_1, H_2), (T_1, T_2)\}$

Introduction to Probability and Combinatorics

- **Event *E***
  Any subset ***E*** of the sample space is known as an event. That is, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in ***E***, then we say that ***E*** has occurred.

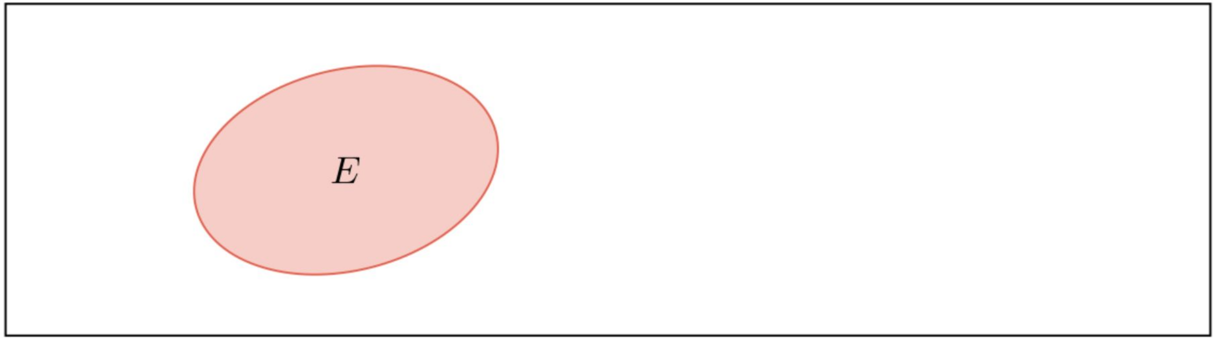
- Examples
  - Getting at least 2 tails in 3 coin toss, *ie*
    $\boldsymbol{E} = \{(H_1, T_2, T_3), (T_1, H_2, T_3), (T_1, T_2, H_3), (T_1, T_2, T_3)\}$
    where
    $\boldsymbol{S} = \{(H_1, H_2, H_3), (H_1, H_2, T_3), (H_1, T_2, H_3), (T_1, H_2, H_3),$
    $(H_1, T_2, T_3), (T_1, H_2, T_3), (T_1, T_2, H_3), (T_1, T_2, T_3)\}$

Introduction to Probability and Combinatorics

- **Axioms of Probability**
  For each event ***E***, we denote ***P(E)*** as the probability of event ***E*** occurring.

- ***Axiom 1*** - Every probability is between 0 and 1 included.

$$\boxed{0 \leqslant P(E) \leqslant 1}$$

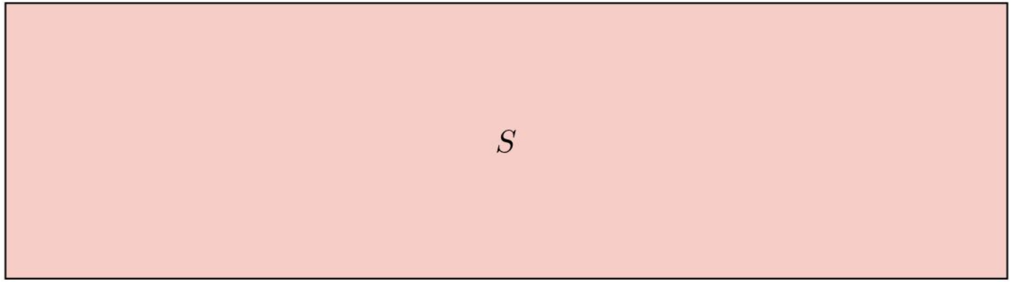


Introduction to Probability and Combinatorics

- **Axioms of Probability**
  For each event ***E***, we denote ***P(E)*** as the probability of event ***E*** occurring.

- ***Axiom 2*** - The probability that at least one of the elementary events in the entire sample space will occur is 1.

$$\boxed{P(S) = 1}$$

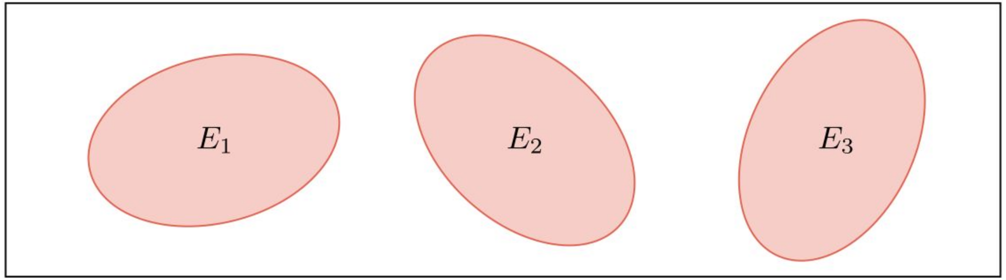


Introduction to Probability and Combinatorics

- **Axioms of Probability**
  For each event ***E***, we denote ***P(E)*** as the probability of event ***E*** occurring.

- ***Axiom 3*** - For any sequence of mutually exclusive events $E_1, \ldots, E_n$, we have:

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i)$$



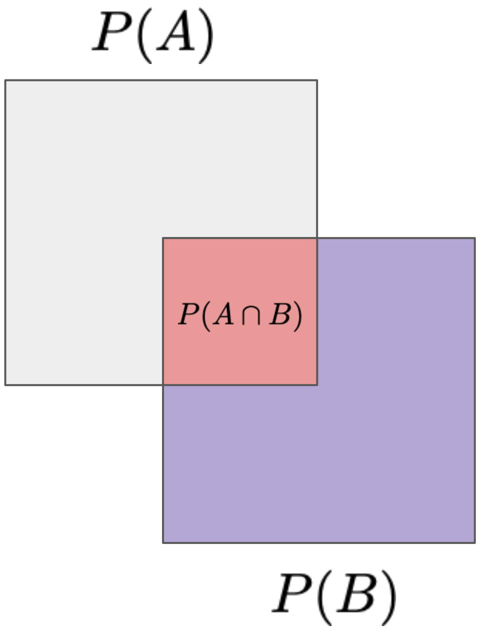
Conditional Probability

- **Conditional Probability**
  For events **A** and **B**; the probability of event **A**, given that event **B** has happened is shown as:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- For events **A** and **B**; the probability of event **A** and event **B** have happened is shown as:

$$P(A \cap B) \quad \text{or} \quad P(A, B)$$



Conditional Probability

- **Chain Rule**
  Intersection of many events can be decomposed into conditional probabilities:

$$P(A_1 \cap A_2 ... \cap A_n) = \prod_{i=1}^{n} P(A_i | \bigcap_{j=1}^{i-1} A_j)$$

- Example

$$\begin{aligned} P(A \cap B \cap C) &= P(A|B \cap C)\ P(B \cap C) \\ &= P(A|B \cap C)\ P(B|C)\ P(C) \end{aligned}$$

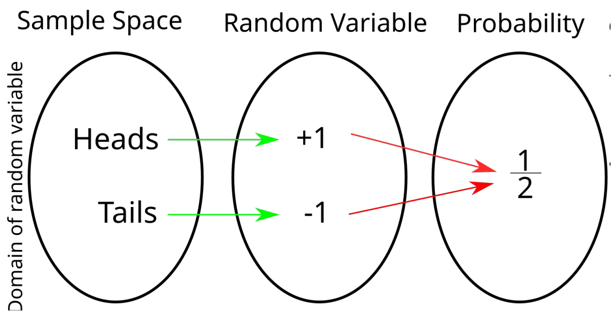
Conditional Probability

- **Bayes Rule**
  For events **A** and **B** such that **P(B)** > 0, **P(A)** > 0, we have:

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

Random Variables

- **Random Variable**
  A random variable, often noted ***X***, is a mathematical formalization of a quantity or object which depends on random events
  - The domain is the set of possible outcomes in a sample space ***S***
  - The range is a measurable space corresponding to the domain



Random Variables

- **Probability Density Function (PDF)**
  The probability density function ***f*** is the probability that ***X*** takes on values between two adjacent realizations of the random variable.

  Can be interpreted as providing a relative likelihood that the value of the ***X***.

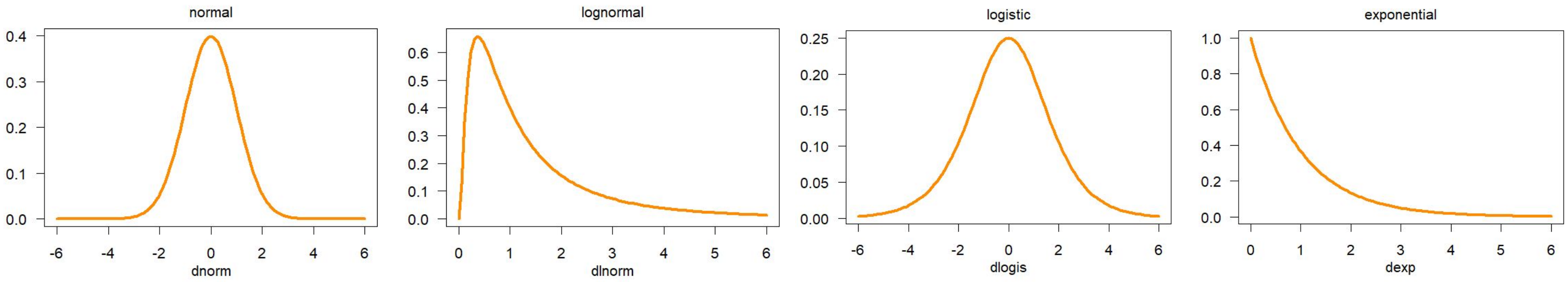


Figure: Some PDFs of different distributions

Random Variables

- **Cumulative Distribution Function (CDF)**
  The cumulative distribution function F, which is monotonically non-decreasing and is such that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$ is defined as:

$$\boxed{F(x) = P(X \leqslant x)}$$

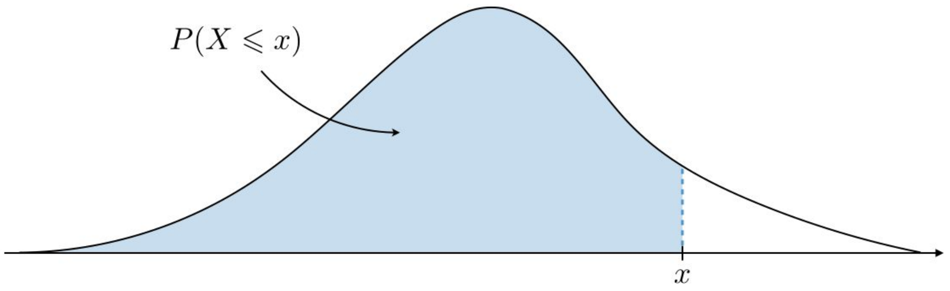


Figure: The definition of CDF over PDF

Random Variables

- **Relations between CDF and PDF**

CDF F	PDF f	Properties of PDF
$F(x) = \int_{-\infty}^{x} f(y)dy$	$f(x) = \frac{dF}{dx}$	$f(x) \geqslant 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$

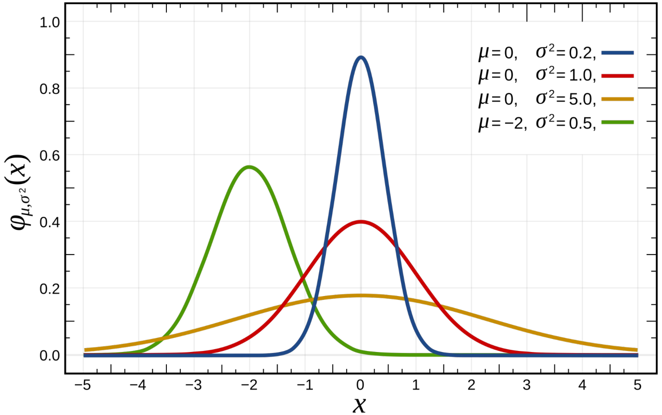


Figure: PDFs of Gaussians

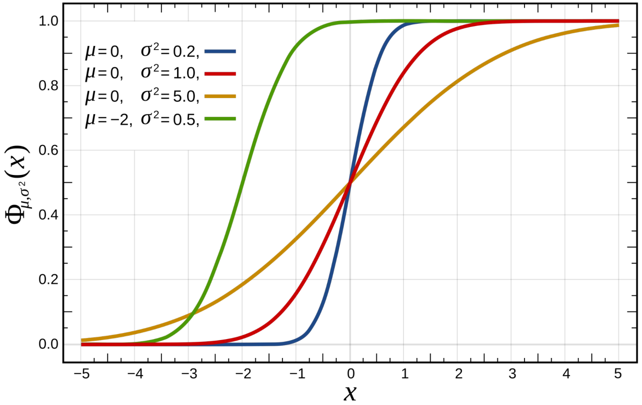


Figure: CDFs of Gaussians

Random Variables

- **Expected Value**
  The expected value ***E[X]*** is the mean of the possible values a random variable can take, weighted by the probability of those outcomes:

$$E[X] = \sum_x x \, P(x)$$

- Example
  Let ***X*** represent the outcome of a roll of a fair six-sided die. The expectation of ***X*** is:

$$\mathrm{E}[X] = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5$$

Random Variables

- **Variance**
  The variance of a random variable, often noted ***Var(X)*** or σ^2 is a measure of the spread of its distribution function. It is determined as follows:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\
&= \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}[X]^2\right] \\
&= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2 \\
&= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]^2 + \mathrm{E}[X]^2 \\
&= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2
\end{aligned}$$

Random Variables

- **Variance and Mean**

  The variance of a collection of ***n*** equally likely values can be written as:

$$\mathrm{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

where μ is the average value, *ie* **mean**. That is,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Random Variables

- **Standard Deviation**
  The standard deviation of a random variable, often noted σ, is a measure of the spread of its distribution function which is compatible with the units of the actual random variable. It is determined as follows:

$$\boxed{\sigma = \sqrt{\mathrm{Var}(X)}}$$

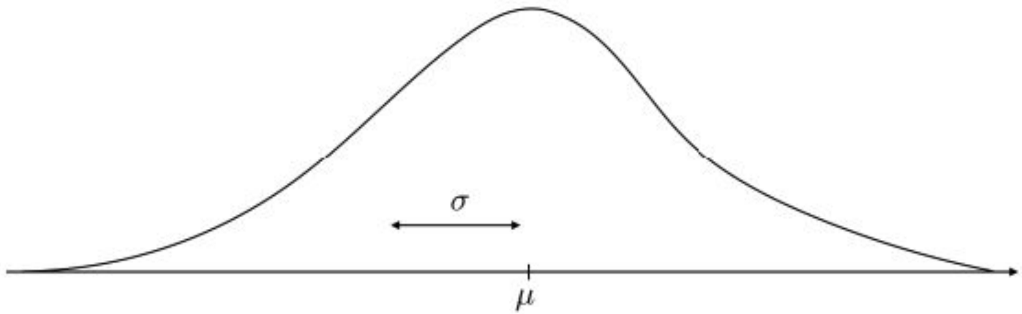


Figure: Standard Deviation of a Gaussian

Probability Distributions

- **Main Distributions: Binomial (Discrete)**

Distribution	PDF	$E[X]$	$\mathrm{Var}(X)$
$X \sim B(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$

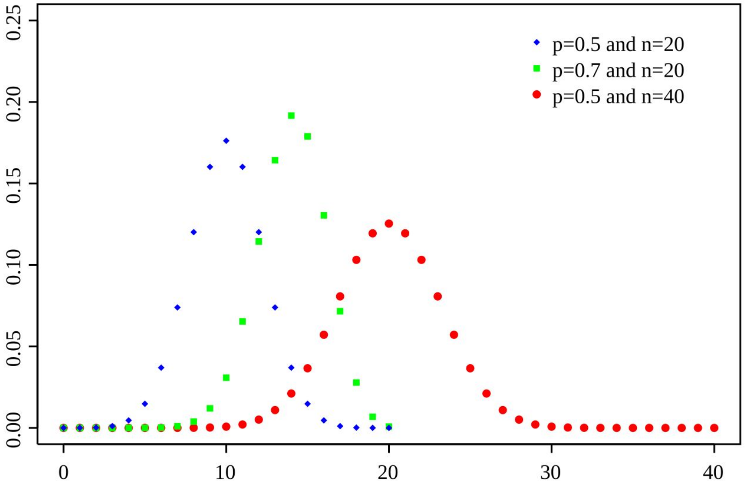


Figure: PDF of Binomial

Probability Distributions

- **Main Distributions: Uniform (Continuous)**

Distribution	PDF	$E[X]$	$\mathrm{Var}(X)$
$X \sim \mathcal{U}(a,b)$	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$

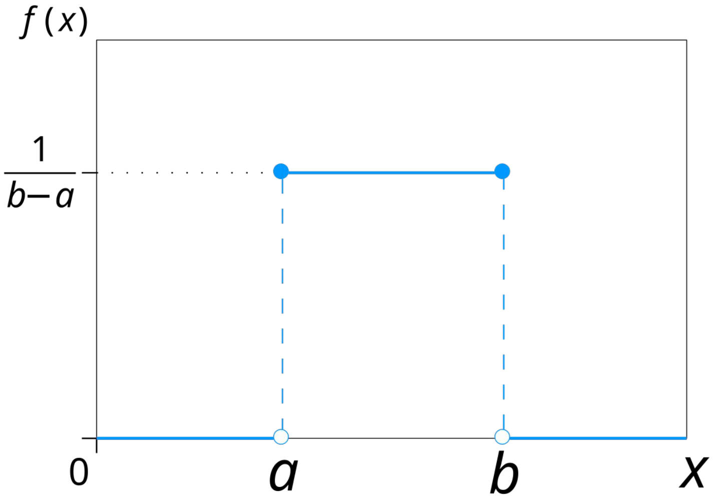


Figure: PDF of Uniform

Probability Distributions

- **Main Distributions: Gaussian (Continuous)**

Distribution	PDF	$E[X]$	$\mathrm{Var}(X)$
$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

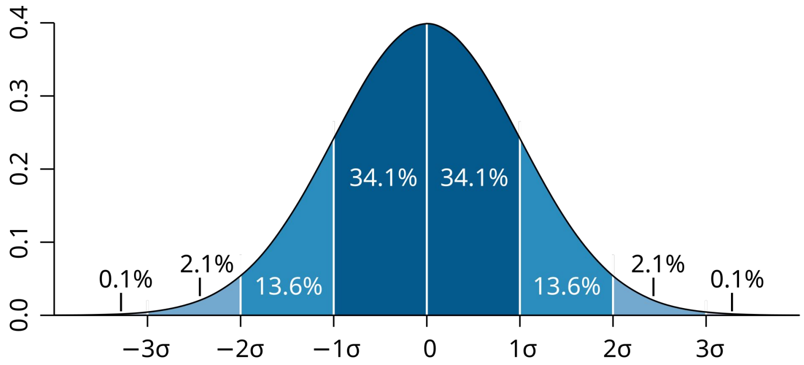


Figure: PDF of Gaussian

Probability Distributions

- **Main Distributions: Multivariate Gaussian (Continuous)**
  The normal distribution generalizes to $\mathbf{R}^n$. It may be parametrized with a positive definite symmetric matrix ∑, ie covariance matrix:

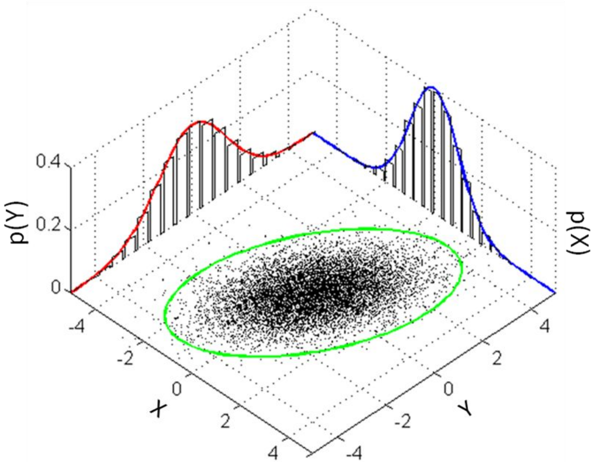


Figure: PDF of Multivariate Gaussian

$$\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sqrt{\frac{1}{(2\pi)^n \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Probability Distributions

- **Main Distributions: Laplace (Continuous)**

Distribution	PDF	$E[X]$	$\mathrm{Var}(X)$
$\mathrm{Laplace}(\mu, b)$	$\frac{1}{2b}\exp\left(-\frac{\lvert x-\mu\rvert}{b}\right)$	μ	$2b^2$

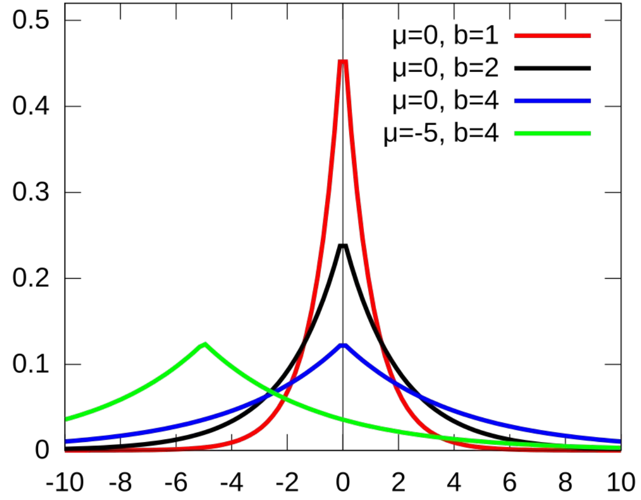


Figure: PDF of Laplace

Probability Distributions

- **Sigmoid Function**
  Certain functions arise often while working with probability distributions, especially the probability distributions used in deep learning models, such as Logistic Sigmoid:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

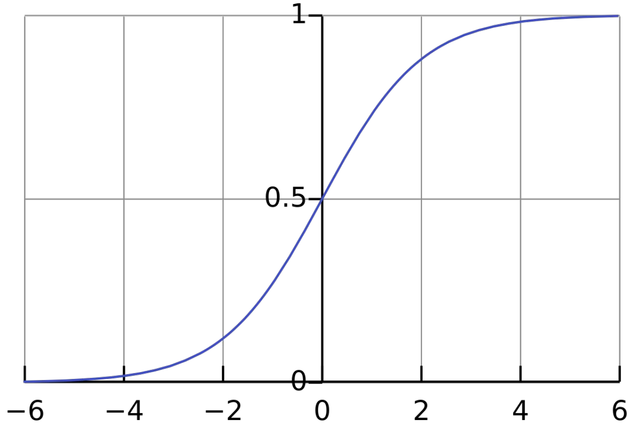


Figure: Logistic curve

Any Questions?