REVIEW ON PROBABILITY

Nesma Refaei
nesma.a.refaei@gmail.com

Machine learning

- Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

- The process of learning begins with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples that we provide.

Machine learning tasks

1. Supervised machine learning (Classification / Regression)
2. unsupervised machine learning (Clustering)
3. Semi-supervised machine learning
4. Reinforcement machine learning

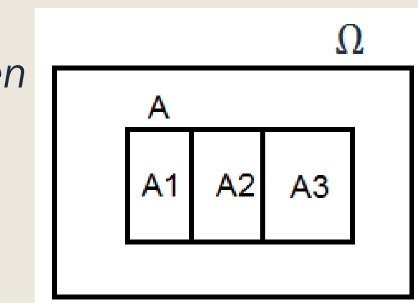
Probability Theory

- Randomness is all around us.
- Probability theory is the study of uncertainty.
- Probability theory is the mathematical framework that allows us to analyze chance events in a logically sound manner.
- The probability of an event is a number indicating how likely that event will occur. This number is always between 0 and 1, where 0 indicates impossibility and 1 indicates certainty.

Elements of probability

- Sample space Ω: The set of all the outcomes of a random experiment.
- Set of events (or event space) F: A set whose elements A ∈ F (called events) are subsets of Ω
- Probability measure: A function $P : F \longrightarrow R$ that satisfies the following properties,
  - *$0 \leq P(A) \leq 1$, for all $A \in F$*
  - *$P(\Omega) = 1$*
  - *$P(\Phi) = 0$*
  - *If A1, A2, . . . are disjoint events (i.e., $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then*

$$P(A) = \sum_i P(A_i)$$



Elements of probability

- Probability measure:
  - $P(A^c) = P(\Omega) - P(A) = 1 - P(A)$
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 1 - P(A)$
  - *Mutual exclusive events:* $P(A \cap B) = 0$
  - *Independent events:* $P(A|B) = P(A)$

Elements of probability

- **Conditional Probability:** Probability of event *A* given event *B*.

  $P(A \text{ given } B) = P(A \mid B)$

- **Joint Probability:** Probability of events *A* and *B*.

  $P(A, B) = P(A \mid B) * P(B)$

- **Marginal Probability:** Probability of event X=*A* given variable *Y*.

  $P(X=A) = \text{sum } P(X=A, Y=yi) \text{ for all } y$

Bayes Rule

Bayes' theorem is stated mathematically as the following equation:[3]

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

where A and B are events and $P(B) \neq 0$.

- $P(A \mid B)$ is a conditional probability: the probability of event A occurring given that B is true. It is also called the posterior probability of A given B.
- $P(B \mid A)$ is also a conditional probability: the probability of event B occurring given that A is true. It can also be interpreted as the likelihood of A given a fixed B because $P(B \mid A) = L(A \mid B)$.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B respectively without any given conditions; they are known as the marginal probability or prior probability.
- A and B must be different events.

Random variables and probability distributions

- A **random variable** is a numerical description of the outcome of a statistical experiment.
- A random variable that may assume only a **finite** number of values is said to be **discrete.**
- One that may assume **any value in some interval** on the real number line is said to be **continuous.**

Random variables and probability distributions

- The **probability distribution** for a random variable describes how the probabilities are distributed over the values of the random variable.

- For **a discrete random variable**, x, the probability distribution is defined by a **probability mass function**, denoted by f(x).
  - *This function provides the probability for each value of the random variable.*
  - *Two conditions must be satisfied: (1) f(x) must be nonnegative for each value of the random variable, and (2) the sum of the probabilities for each value of the random variable must equal one.*

Random variables and probability distributions

- For **a continuous random variable**, x, it is not meaningful to talk about the probability that the random variable will take on a specific value; instead, the probability that a continuous random variable will lie within a given interval is considered. This is called the **probability density function,** also denoted by $f(x)$.
- The probability that the variable will take on a value within an interval is the area under the graph of $f(x)$ corresponding to that interval, obtained by computing the integral of $f(x)$ over that interval. A probability density function must satisfy two requirements: (1) $f(x)$ must be nonnegative for each value of the random variable, and (2) the integral over all values of the random variable must equal one.

Expectation

- The **expected value**, or **mean**, of a random variable—denoted by $E(x)$ or μ—is a weighted average of the values the random variable may assume.
- In the discrete case the weights are given by the probability mass function, and in the continuous case the weights are given by the probability density function.
- $E(x) = \Sigma x f(x)$ (For discrete random variables)
- $E(x) = \int x f(x) dx$ (For continuous random variables)

Variance

- The **variance** of a random variable, denoted by Var(x) or σ^2, is a weighted average of the squared deviations from the mean.
- Var(x) = E($(x-\mu)^2$)

$$\mathrm{Var}(x) = \sigma^2 = \Sigma(x-\mu)^2 f(x)$$

$$\mathrm{Var}(x) = \sigma^2 = \int (x-\mu)^2 f(x)dx$$

- **Standard deviation** denoted by σ is the squared root of Var(x)

Properties

- $E(a x + b) = a E(x) + b$
- $E(x + y) = E(x) + E(y)$
- $E(g(x)) = \Sigma g(x) f(x)$ or $E(g(x)) = \int g(x) f(x)$
- $Var(a x + b) = a^2 Var(x)$

Special probability distributions

- Continuous uniform distribution

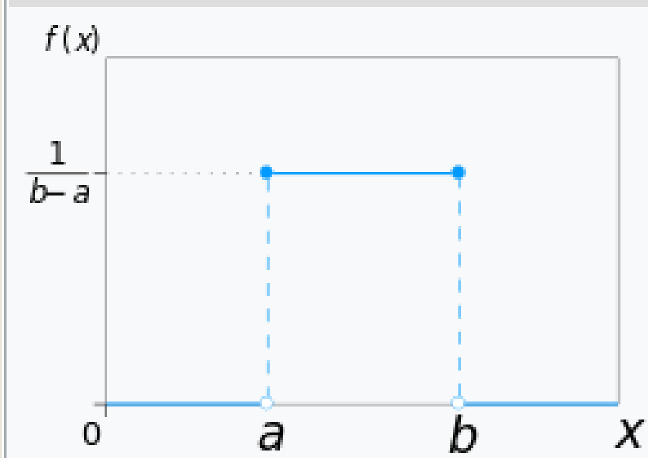
The probability density function of the continuous uniform distribution is:

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

$$E(X) = \frac{1}{2}(b+a).$$

$$V(X) = \frac{1}{12}(b-a)^2$$

Probability density function



Using maximum convention

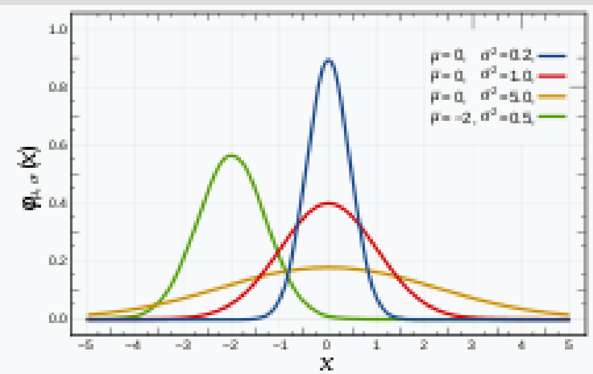
Special probability distributions

- Normal (Gaussian) distribution

The general form of its probability density function is

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

Probability density function



The red curve is the *standard normal distribution*

LET’S PRACTICE

Question 1

1) Consider the numbers -2,3,0.5,0.3,4.2,9,-3.4.

Find the mean, the variance and the standard deviation. If the points are multiplied a constant a, what will be the mean, the variance and the standard deviation? If we add to the above points the point 10,000. Without explicitly computing the standard deviation, will it be higher or lower?

Question 2

- If $E[X] = 1$ and $Var(X) = 5$, find:

(a) $E[(2 + X)^2]$.

(b) $Var(4 - 3X)$.

Question 3

There are 5 boxes of mixed fruits with 1 of the boxes is red, 1 is blue and the rest are green. Each box color has a mixture of fruits according to the table:

Apply the **<u>sum and product rules</u>** to compute the following:
a) Compute $p(\text{Fruit} = \textit{Apple})$.
b) Compute $p(\text{Fruit} = \textit{Orange} \mid \text{Box} = \textit{Red})$.
c) Compute $p(\text{Box} = \textit{Green} \mid \text{Fruit} = \textit{Orange})$.

	Red Box	Blue Box	Green Box
Apples	3	5	3
Oranges	4	5	3
Limes	3	0	4

Question 4

- The amount of time in hours that a computer functions before breaking down is a continuous random variable with probability density function given by

- $f(x) = \begin{cases} \lambda e^{-\frac{x}{100}} & x \geq 0 \\ 0 & x < 0 \end{cases}$

- What is the probability that:

(a) a computer will function between 50 and 150 hours before breaking down?

(b) it will function for fewer than 100 hours?