

# Artificial Intelligence

Arland Barrera

November 7, 2025

---

# Contents

---

<b>1</b>	<b>Statistics</b>	<b>5</b>
1.1	Basic Concepts . . . . .	5
1.1.1	Mean . . . . .	5
1.1.2	Weighted Average . . . . .	6
1.1.3	Median . . . . .	6
1.1.4	Mode . . . . .	6
1.1.5	Range . . . . .	6
1.1.6	Variance . . . . .	7
1.1.7	Standard Deviation . . . . .	7
1.1.8	Gaussian Distribution . . . . .	8
1.2	Metrics . . . . .	9
1.2.1	Accuracy . . . . .	9
1.2.2	Recall . . . . .	9
1.2.3	Precision . . . . .	9
1.2.4	F1-Score . . . . .	10
<b>2</b>	<b>Mathematics</b>	<b>11</b>
2.1	Some Section . . . . .	11
2.1.1	Sigmoid . . . . .	11
<b>3</b>	<b>Algorithms</b>	<b>12</b>
3.1	Supervised Learning . . . . .	12
3.1.1	Linear Regression . . . . .	12
3.1.2	Logistic Regression . . . . .	12
3.1.3	K-Nearest Neighbor . . . . .	12
3.1.4	Decision Tree . . . . .	12
3.1.5	Support Vector Machine . . . . .	12
3.1.6	Naive Bayes . . . . .	12
3.2	Unsupervised Learning . . . . .	13
3.3	Optimization . . . . .	13

---

# List of Plots

---

1.1 Gaussian distribution . . . . . 9

2.1 Sigmoid function . . . . . 11

---

# List of Equations

---

1.1	Sample mean	5
1.2	Population Mean	5
1.3	Weighted mean	6
1.4	Odd median	6
1.5	Even median	6
1.6	Range	7
1.7	Sample variance	7
1.8	Population variance	7
1.9	Sample standard deviation	8
1.10	Population standard deviation	8
1.11	Probability density function	8
1.12	Accuracy	9
1.13	Recall	9
1.14	Precision	10
1.15	F1-Score	10
2.1	Sigmoid function	11
3.1	Naive Bayes theorem	13

---

# Statistics

---

Statistics is a field of mathematics that pertains to data analysis. Statistical methods and equations can be applied to a data set in order to analyze and interpret results, explain variations in the data, or predict future data.

Two important concepts in Statistics are **population** and **sample**. Population is all the data available, the larger group. Sample is a subset of the population, a fraction or portion of the total data available.

## 1.1 Basic Concepts

### 1.1.1 Mean

The mean, also known as the **arithmetic mean** or **average**, is obtained by dividing the sum of observed values, data values, by the number of observations. This gives a good idea as to what the points are closest to. Occasionally a value that differs greatly can be seen, the mean would incorporate the occasional outlying data. This means the average is susceptible to outlying values. The idea is to find the center of the data set. It is the most common type of average.

The **sample mean** is represented by an  $\bar{x}$  with an over line. The number of samples is represented by the letter  $n$ .

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.1)$$

Equation 1.1: Sample mean

The **population mean** is represented by the greek letter mu,  $\mu$ . The total number of values is represented by the letter  $N$ .

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (1.2)$$

Equation 1.2: Population Mean

Both equations 1.1 and 1.2 are the same, the only change are the symbols employed. They can be used when the error associated with each measurement is the same or unknown.

### 1.1.2 Weighted Average

Also known as **weighted mean** or **weighted arithmetic mean**. The weighted average gives more importance to certain data points by multiplying each value a specific weight before summing them up. The result is then divided by the sum of the weights to find the average. This method is used when data have different levels of importance or significance.

$$\bar{X}_w = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (1.3)$$

Equation 1.3: Weighted mean

### 1.1.3 Median

The median is the middle value of a set of data containing an odd number of values, or the average of the two middle values of a set of data with an even number of values. The median is specially helpful when separating data into two equal sized distributions. It is useful if there is an interest in knowing the range of values the system could be operating in. Half the values should be above and half the values should be below, so there is an idea of where the middle operating point is. Because of this the median is less sensitive to outlier data.

For **odd numbers** the median is the data points  $n$  plus 1 divided by 2.

$$\text{Median} = \frac{n+1}{2} \quad (1.4)$$

Equation 1.4: Odd median

For **even numbers** the median is the average of the data points  $n$  divided by two, and  $n$  divided by two plus one.

$$\text{Median} = \frac{1}{2} \left( \frac{n}{2} + \left( \frac{n}{2} + 1 \right) \right) \quad (1.5)$$

Equation 1.5: Even median

### 1.1.4 Mode

The mode of a set of data is the value which occurs most frequently.

### 1.1.5 Range

The range is the difference between largest and smallest values. This is a way to find the dispersion of the data set. However, if there are outliers in the data the range is of little use.

$$\boxed{\text{Range} = x_{\max} - x_{\min}} \quad (1.6)$$

Equation 1.6: Range

### 1.1.6 Variance

Measure of dispersion that indicates how far each individual data point is from the mean. This is the spread of each individual data point.

The deviation of each data point from the mean is, the data value minus the mean:

$$x_i - \bar{X}$$

Adding up each deviation and then divide by the number of data points, averages the deviation of every point from the mean. This is the average variance or spread of the data from the mean.

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})$$

This has a problem. The values to the left of the mean result in negative values, and the ones to the right are positive. To solve this, the deviation of each data point is squared. This results in the **variance**,

The **sample variance** is represented by the letter  $s$  squared,  $s^2$ . This is divided by the sample data minus 1,  $n - 1$ .

$$\boxed{s^2 = \frac{1}{n - 1} \sum_{i=1}^n (x_i - \bar{X})^2} \quad (1.7)$$

Equation 1.7: Sample variance

The **population variance** is represented by the greek letter sigma to the power of two,  $\sigma^2$ . This is divided by the total population data,  $N$ .

$$\boxed{\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1.8)$$

Equation 1.8: Population variance

### 1.1.7 Standard Deviation

The variance squares the spread of each data point, this raises a problem. The measure of units of the data set is lost because it has been squared. To solve this problem the square root of the variance can be found, that way the measure of the original data is more close. This is known as standard deviation.

The standard deviation gives an idea of how close the entire set of data is to the average value. Data sets with a small standard deviation have tightly grouped, precise data. Data sets with large standard deviations have data spread out over a wide range of values.

The **sample standard deviation** is represented by the letter  $s$ .

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{X})^2} \quad (1.9)$$

Equation 1.9: Sample standard deviation

The **population standard deviation** is represented by the greek letter  $\sigma$ .

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1.10)$$

Equation 1.10: Population standard deviation

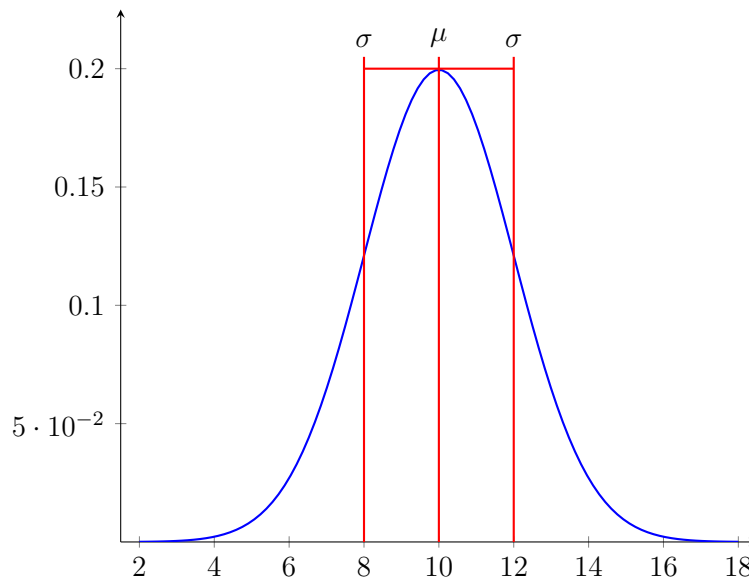
### 1.1.8 Gaussian Distribution

Gaussian distribution, also known as **normal distribution**, is represented by the following **Probability Density Function**:

$$\text{PDF}_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1.11)$$

Equation 1.11: Probability density function

where  $\mu$  is the mean and  $\sigma$  is the standard deviation of a very large data set. The Gaussian distribution is a bell-shaped curve, symmetric about the mean value. Probability density functions represent the spread of a data set. An example is shown below.



Plot 1.1: Gaussian distribution

In this specific example,  $\mu = 10$  and  $\sigma = 2$ .

Normal distribution is completely defined by the mean ( $\mu$ ) and the standard deviation ( $\sigma$ ). The line never touches the x-axis. The y-axis acts as the probability. Total area under curve is 1, because the sum of all probabilities is 1.

## 1.2 Metrics

### 1.2.1 Accuracy

Proportion of correct predictions and total predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1.12)$$

Equation 1.12: Accuracy

### 1.2.2 Recall

Proportion of true positives and total positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1.13)$$

Equation 1.13: Recall

### 1.2.3 Precision

Proportion of true positives and positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1.14)$$

Equation 1.14: Precision

#### 1.2.4 F1-Score

Mean of precision and recall.

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (1.15)$$

Equation 1.15: F1-Score

## 2.1 Some Section

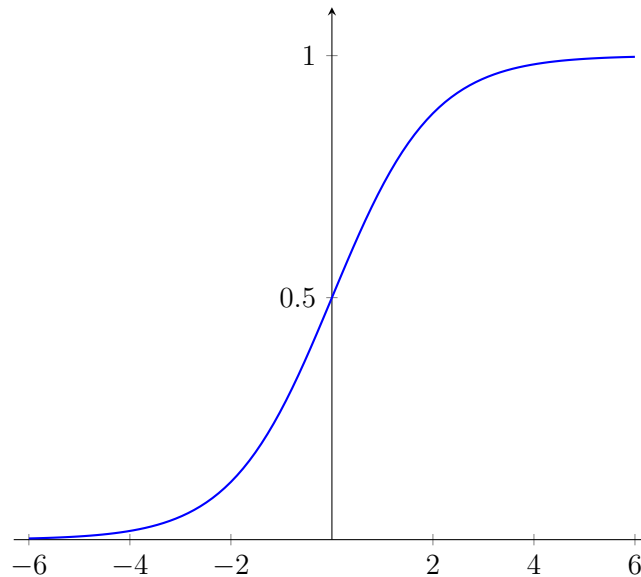
### 2.1.1 Sigmoid

This function takes any real-value input and maps it to an output between 0 and 1, creating a characteristic S-shaped curve. It is also known as **logistic curve**.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.1)$$

Equation 2.1: Sigmoid function

The S-shaped curve of the function is shown below.



Plot 2.1: Sigmoid function

#### Characteristics:

- Domain:  $(-\infty, \infty)$
- Range:  $(0, +1)$
- Intersection:  $f(0) = 0.5$

---

# Algorithms

---

Artificial intelligence algorithms can be classified by their learning approach, task type, domain, and others.

## Learning approach

- Supervised learning.
- Unsupervised learning.
- Reinforcement learning.
- Semi-supervised learning.

## Task type

- Classification: used to sort data into categories.
  - Binary classification: separates data into two classes.
  - Multi-class classification: assigns an item to three or more classes.
- Regression: used to predict a continuous value.

## Domain

- Natural Language Procesing (NLP).
- Computer vision.

## 3.1 Supervised Learning

### 3.1.1 Linear Regression

### 3.1.2 Logistic Regression

### 3.1.3 K-Nearest Neighbor

### 3.1.4 Decision Tree

### 3.1.5 Support Vector Machine

### 3.1.6 Naive Bayes

The Naive Bayes algorithm in machine learning is a probabilistic classification algorithm based on Bayes' theorem that makes a "*naive*" assumption that all features are independent of each other given

the class. It calculates the probability of a class given a set of features, using the formula shown below.

$$P(\text{class}|\text{data}) = \frac{P(\text{data}|\text{class})P(\text{class})}{P(\text{data})} \quad (3.1)$$

Equation 3.1: Naive Bayes theorem

The independence assumption simplifies the calculation.

## 3.2 Unsupervised Learning

## 3.3 Optimization