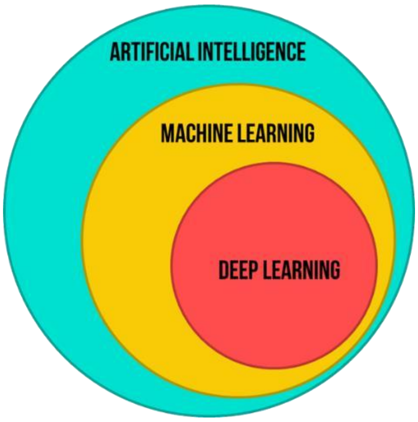
Introduction to Machine Learning

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

Example: Image recognition, Speech recognition, Medical diagnosis, Statistical arbitrage, Predictive analytics, etc.

Artificial Intelligence, Machine Learning and Deep Learning

- **Artificial Intelligence** is defined as a program that exhibits cognitive ability similar to that of a human being. It makes computers think like humans and solve problems the way we do is one of the main tenets of artificial intelligence.
- Any computer program that shows characteristics, such as self-improvement, learning through inference, or even basic human tasks, such as image recognition and language processing, is considered to be a form of AI.
- The field of artificial intelligence includes within it the sub-fields of machine learning and deep learning.
- **Deep Learning** is a more specialized version of machine learning that utilizes more complex methods for difficult problems.


ARTIFICIAL INTELLIGENCE
MACHINE LEARNING
DEEP LEARNING

Definition of machine learning

Arthur Samuel, an early American leader in the field of computer gaming and artificial intelligence, coined the term “Machine Learning” in 1959 while at IBM.

He defined machine learning as “the field of study that gives computers the ability to learn without being explicitly programmed.” However, there is no universally accepted definition for machine learning. Different authors define the term differently. We give below two more definitions.

1. Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data, or both.

2. The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve with experience

Definition of learning

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks T, as measured by P, improves with experience E.

Examples :

i) Handwriting recognition learning problem

• Task T: Recognising and classifying handwritten words within images

• Performance P: Percent of words correctly classified

• Training experience E: A dataset of handwritten words with given classifications

ii) A robot driving learning problem

• Task T: Driving on highways using vision sensors

• Performance measure P: Average distance traveled before an error

• training experience: A sequence of images and steering commands recorded while observinga human driver

iii) A chess learning problem

• Task T: Playing chess

• Performance measure P: Percent of games won against opponents

• Training experience E: Playing practice games against itself

Definition: A computer program which learns from experience is called a machine learning program or simply a learning program. Such a program is sometimes also referred to as a learner.

Basic components of learning process

The learning process, whether by a human or a machine, can be divided into four components, namely, data storage, abstraction, generalization and evaluation. Figure 1.1 illustrates the various components and the steps involved in the learning process.

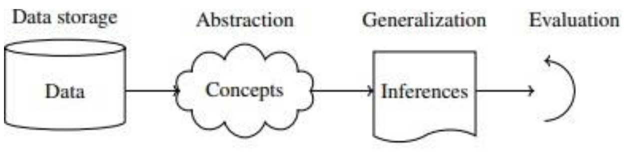


Figure 1.1: Components of learning process

1. Data storage

Facilities for storing and retrieving huge amounts of data are an important component of the learning process. Humans and computers alike utilize data storage as a foundation for advancedreasoning.

• In a human being, the data is stored in the brain and data is retrieved using electrochemical signals.

• Computers use hard disk drives, flash memory, random access memory and similar devices to store data and use cables and other technology to retrieve data.

2. Abstraction

The second component of the learning process is known as abstraction. Abstraction is the process of extracting knowledge about stored data. This involves creating general concepts about the data as a whole. The creation of knowledge involves application of known models and creation of new models. The process of fitting a model to a dataset is known as training. When the model has been trained, the data is transformed into an abstract form that summarizes the original information.

3. Generalization

The third component of the learning process is known as generalisation. The term generalization describes the process of turning the knowledge about stored data into a form that can be utilized for future action. These actions are to be carried out on tasks that are similar, but not identical, to those what have been seen before. In generalization, the goal is to discover those properties of the data that will be most relevant to future tasks.

4. Evaluation

Evaluation is the last component of the learning process. It is the process of giving feedback to the user to measure the utility of the learned knowledge. This feedback is then utilised to effect improvements in the whole learning process.

Applications of machine learning

The following is a list of some of the typical applications of machine learning.

1. In retail business, machine learning is used to study consumer behaviour.

2. In finance, banks analyze their past data to build models to use in credit applications, fraud detection, and the stock market.

3. In manufacturing, learning models are used for optimization, control, and troubleshooting.

4. In medicine, learning programs are used for medical diagnosis.

5. In telecommunications, call patterns are analyzed for network optimization and maximizing the quality of service.

6. In science, large amounts of data in physics, astronomy, and biology can only be analyzed fast enough by computers. The World Wide Web is huge; it is constantly growing and searching for relevant information cannot be done manually.
7. In artificial intelligence, it is used to teach a system to learn and adapt to changes so that the system designer need not foresee and provide solutions for all possible situations.
8. It is used to find solutions to many problems in vision, speech recognition, and robotics.
9. Machine learning methods are applied in the design of computer-controlled vehicles to steercorrectly when driving on a variety of roads.
10. Machine learning methods have been used to develop programmes for playing games suchas chess, backgammon and Go.

Statistics vs Machine Learning

The major difference between machine learning and statistics is their purpose. Machine learning models are designed to make the most accurate predictions possible. Statistical models are designed for inference about the relationships between variables

1. Machine Learning is an algorithm that can learn from data without relying on rules- based programming.

   Statistical modeling is a formalization of relationships between variables in the data inthe form of mathematical equations.
2. Machine learning is all about predictions, supervised learning, unsupervised learning, etc.

   Statistics is about sample, population, hypothesis, etc.
3. Machine learning is a subfield of computer science and artificial intelligence. It deals with building systems that can learn from data, instead of explicitly programmed instructions.

A statistical model, on the other hand, is a subfield of mathematics.

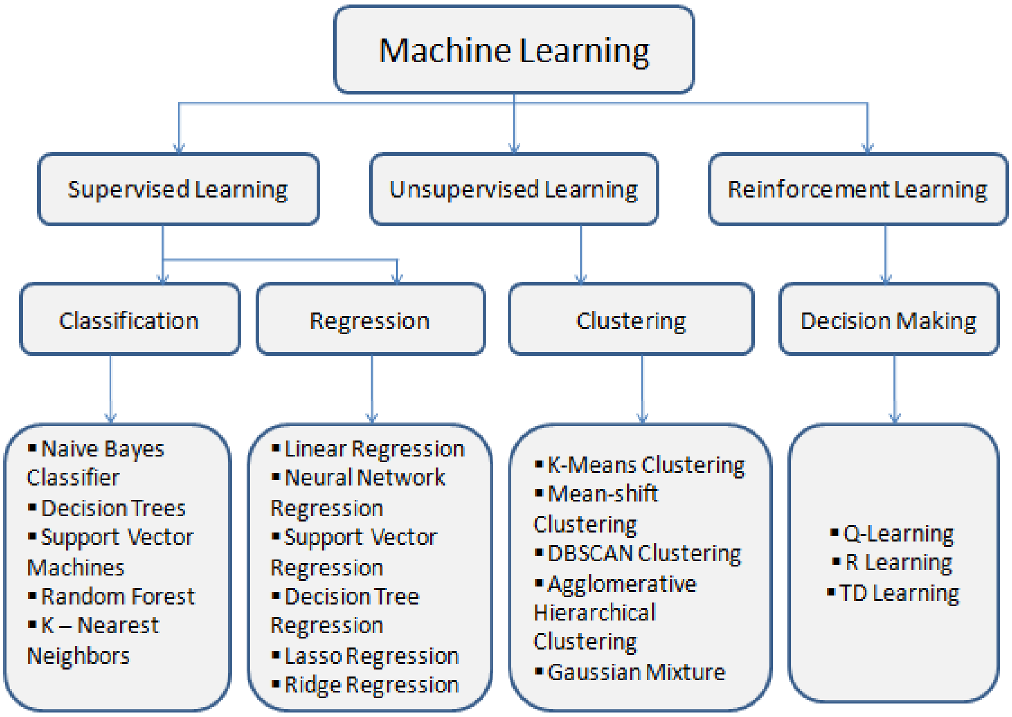
4. Machine Learning is automated and requires less human intervention and it deals withlarge datasets.

   Statistics require a lot of human effort and deals with small datasets.

Types of Machine Learning Algorithms:

These are three types of machine learning:

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning



Supervised Learning:

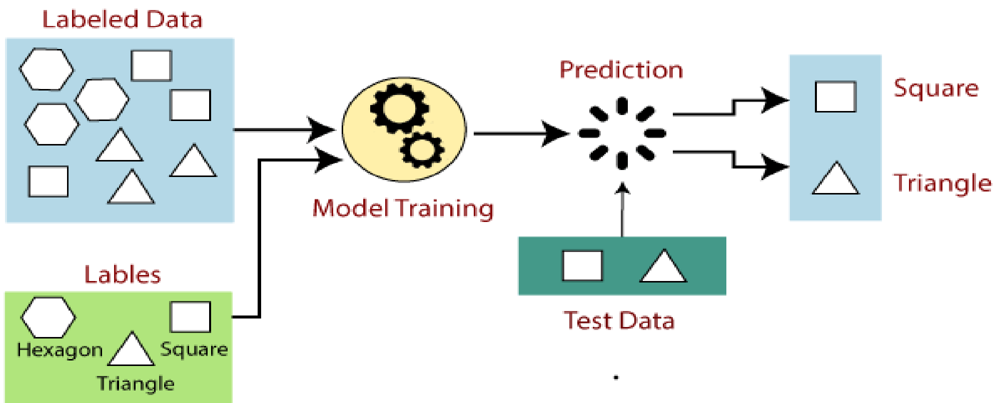
- Supervised learning is one of the most basic types of machine learning.
- In this type, the machine learning algorithm is trained on **labelled data**.
- In supervised learning, the ML algorithm is given **a small training dataset to work with.**
- This training dataset is a **smaller part of the bigger dataset** and serves to give the algorithm a basic idea of the problem, solution, and data points to be dealt with.
- At the end of the training, the algorithm has an idea of how the data

works and therelationship between the input and the output.

- Example: **Risk Assessment, Image classification, Fraud Detection, spam filtering, etc.**

How Supervised Learning Works?

In supervised learning, models are trained using labelled dataset, where the model learns abouteach type of data. Once the training process is completed, the



model is tested on the basis of test data (a subset of the training set), and then it predicts the output.

Suppose we have a dataset of different types of shapes which includes square, rectangle, triangle, and Polygon. Now the **first step is that we need to train the model for each shape.**

- If the given shape has four sides, and all the sides are equal, then it will be labelled asa Square.
- If the given shape has three sides, then it will be labelled as a triangle.
- If the given shape has six equal sides, then it will be labelled as hexagon.

Now, **after training, we test our model using the test set**, and the **task of the model is to identify the shape.** The **machine is already trained on all types of shapes**, and when it finds a new shape, it classifies the shape on the bases of a number of sides, and predicts the output.

Advantages of Supervised learning:

- With the help of supervised learning, the model can predict the output on the basis ofprior experiences.
- In supervised learning, we can have an exact idea about the classes of objects.

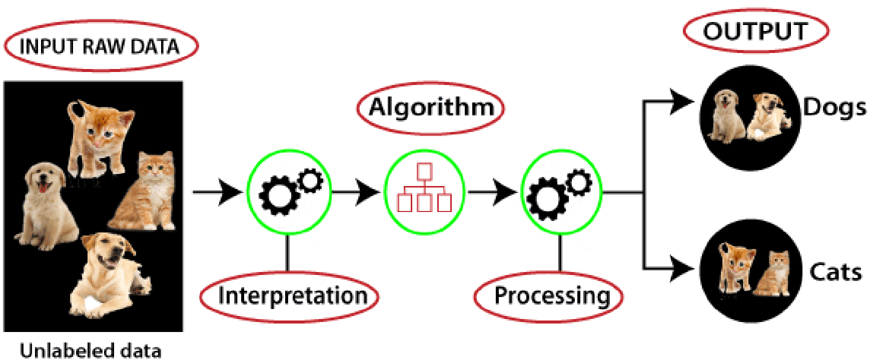
Disadvantages of supervised learning:

- Supervised learning models are not suitable for handling the complex tasks.
- Supervised learning cannot predict the correct output if the test data is different fromthe training dataset.
- Training required lots of computation times.

Unsupervised Learning:

- Unsupervised machine learning holds the advantage of being able to work with **unlabelled data**.
- This means that human labour is not required to make the dataset machine-readable, **allowing much larger datasets to be worked on by the program**.
- This offers more post-deployment development than supervised learning algorithms.
- Example : **Principal Component Analysis, Clustering**

How Unsupervised Learning Works?



Here, we have **taken an unlabelled input data, which means it is not categorized and corresponding outputs are also not given.**

Now, this unlabelled input data is fed to the machine learning model in order to train it.

Firstly, it **will interpret the raw data** to find the hidden patterns from the data and then will apply suitable algorithms such as **k-means clustering, Decision tree, etc.**

Once it applies the suitable algorithm, the **algorithm divides the data objects into groups according to the similarities and difference between the objects.**

Advantages of Unsupervised Learning

- Unsupervised learning is used for more complex tasks as compared to supervised learning because, in unsupervised learning, we don't have labelled input data.
- Unsupervised learning is preferable as it is easy to get unlabelled data in comparison to labelled data.

Disadvantages of Unsupervised Learning

- Unsupervised learning is intrinsically more difficult than supervised learning as it does not have corresponding output.

- The result of the unsupervised learning algorithm might be less accurate as input data isnot labelled, and algorithms do not know the exact output in advance.

Semi-Supervised learning

Semi-supervised learning **bridges supervised learning and unsupervised learning techniques** to solve their key challenges. With it, you train an initial model on a few labeled samples and then iteratively apply it to the greater number of unlabeled data.

- SSL works for a variety of problems from classification and regression to clustering and association.
- uses small amounts of labeled data and also large amounts of unlabeled data, which reduces expenses on manual annotation and cuts data preparation time.

Working of Semi-Supervised Learning

Semi-supervised learning uses pseudo labeling to train the model with less labeled training data than supervised learning. The process can combine various neural network models and trainingways.

- **Firstly, it trains the model with less amount of training data similar to the supervisedlearning models. The training continues until the model gives**

accurate results.

- The input data in labeled training data and unlabeled training data are also linked.
- In the end, again train the model with the new

combined input .

- It will reduce errors and improve the accuracy of

the model.

Semi-supervised learning models applications

- o Speech Analysis
- o Web content classification
- o Text document classifier

<u>Reinforcement Learning</u>

It is a part of ML where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from those actions.

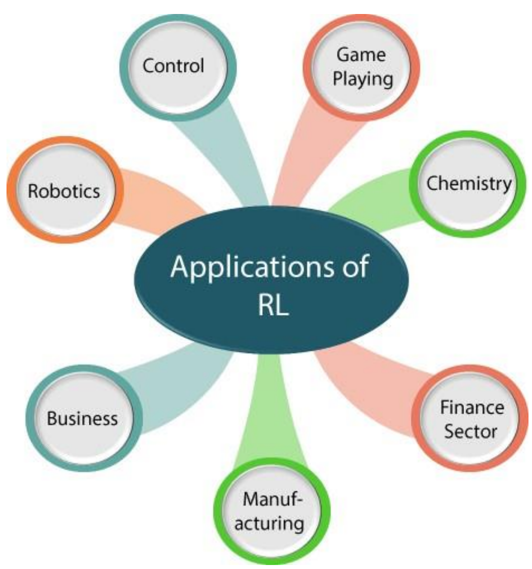
Favourable outputs are encouraged or ‘reinforced’, and non-favourable outputs are discouraged or ‘punished’.

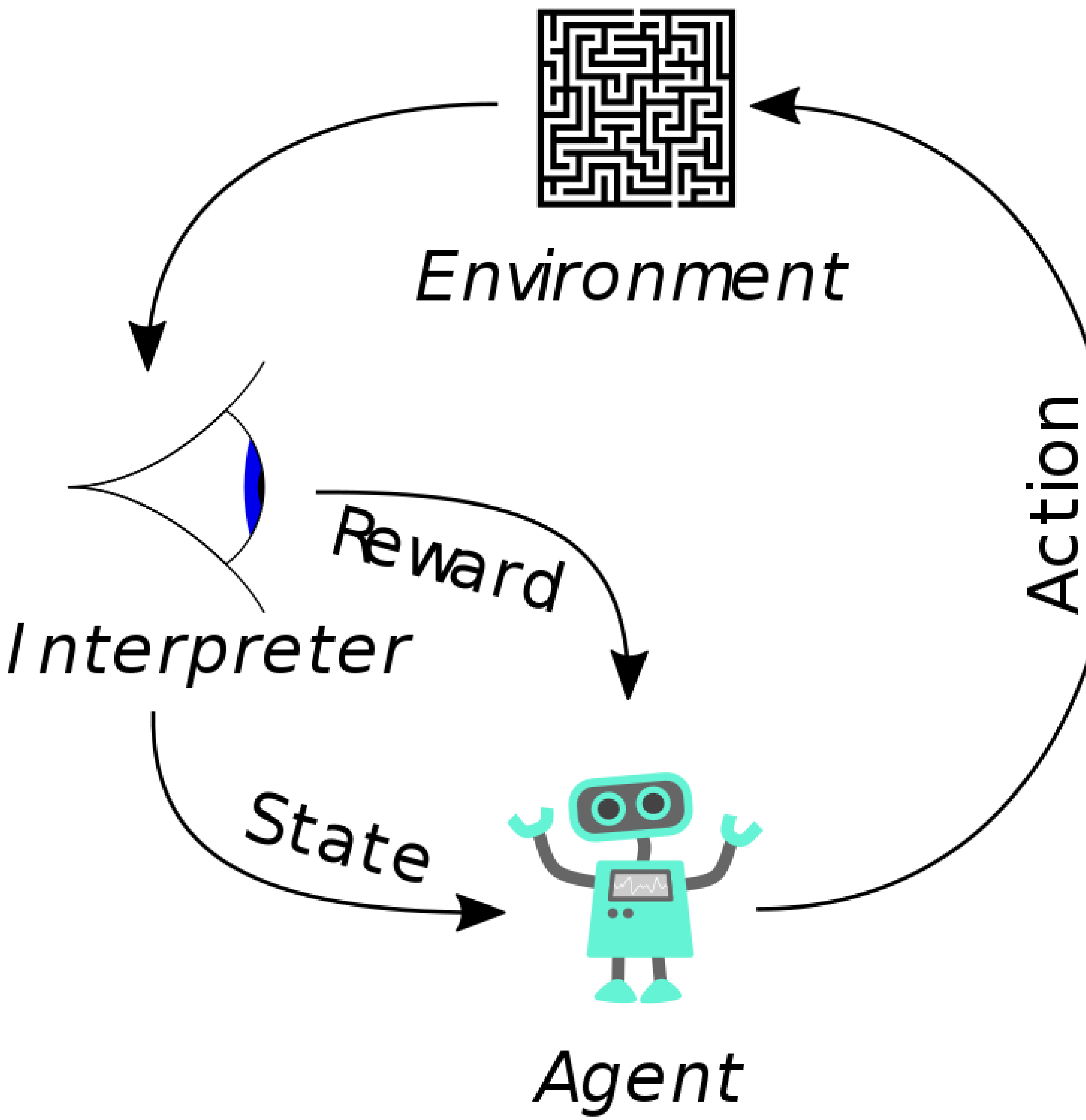
- In every iteration of the algorithm, the output result is given to the interpreter, which decides whether the outcome is favourable or not.
- In case of the **program finding the correct**

solution, the interpreter reinforces the solution by **providing a reward to the algorithm**.

- If the outcome is not favourable, the algorithm is forced to reiterate until it finds a better result.
- In most cases, the reward system is directly tied to the effectiveness of the result.
- In typical reinforcement learning use-cases, such as **finding the shortest route between two points on a map**. The higher this percentage value is, the more reward is given to the algorithm.
- Thus, the program is trained to give the best possible solution for the best possible reward.

Reinforcement Learning Applications





Environment
Interpreter
Reward
State
Action
Agent

Difference Between Supervised, Unsupervised and Reinforcement Learning

Criteria	Supervised ML	Unsupervised ML	Reinforcement ML
Definition	Learns by using labelled data	Trained using unlabelled data without any guidance.	Works on interacting with the environment
Type of data	Labelled data	Unlabelled data	No – predefined data
Type of problems	Regression and classification	Association and Clustering	Exploitation or Exploration
Supervision	Extra supervision	No supervision	No supervision
Algorithms	Linear Regression, Logistic Regression, SVM, KNN etc.	K – Means, C – Means, Apriori	Q – Learning, SARSA
Aim	Calculate outcomes	Discover underlying patterns	Learn a series of action
Application	Risk Evaluation, Forecast Sales	Recommendation System, Anomaly Detection	Self Driving Cars, Gaming, Healthcare

Types of Reinforcement learning

There are mainly two types of reinforcement learning, which are:

- **Positive Reinforcement**

The positive reinforcement learning means adding something to increase the tendency that expected behaviour would occur again. It impacts positively on the behaviour of the agent and increases the strength of the behaviour. This type of reinforcement can sustain the changes for a long time, but too much positive reinforcement may lead to an overload of states that can reduce the consequences.

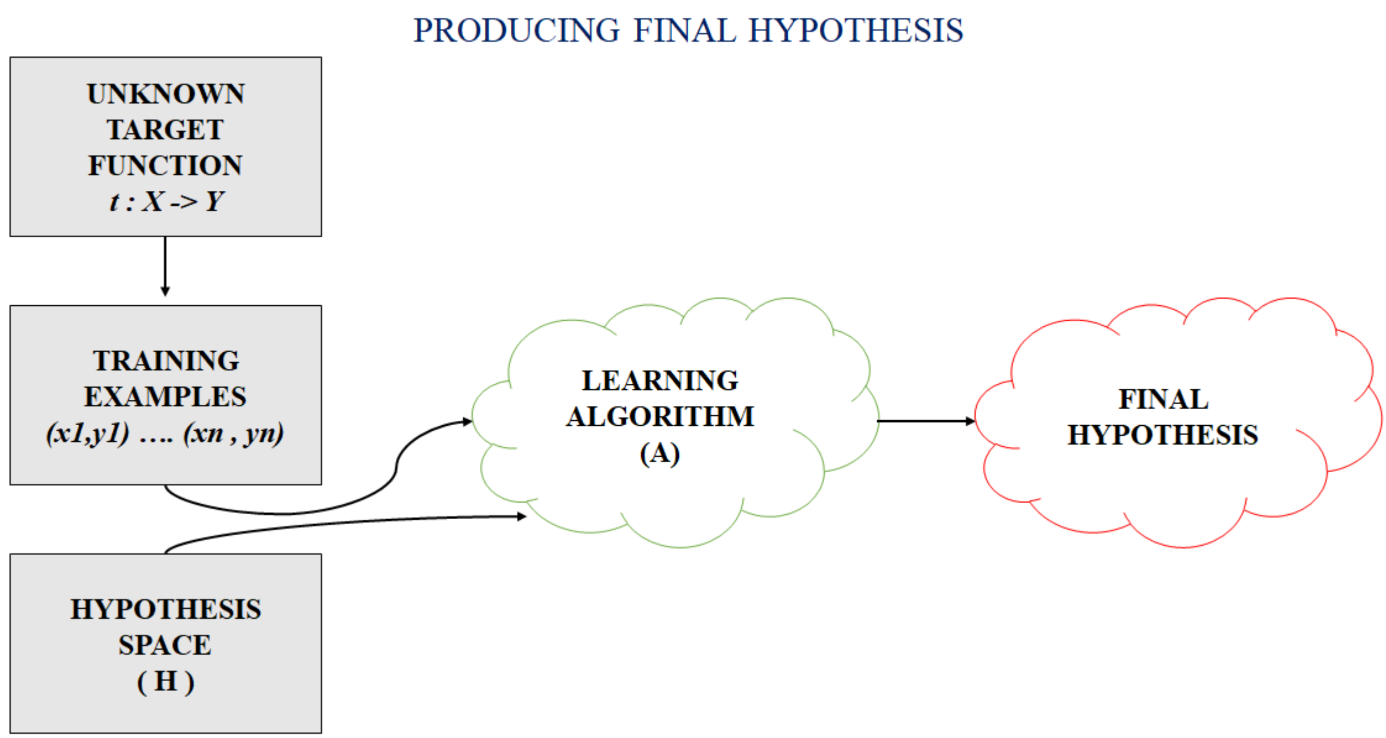
- **Negative Reinforcement:**

The negative reinforcement learning is opposite to the positive reinforcement as it increases the tendency that the specific behaviour will occur again by avoiding the negative

condition. It can be more effective than the positive reinforcement depending on situation and behaviour, but it provides reinforcement only to meet minimum behaviour.

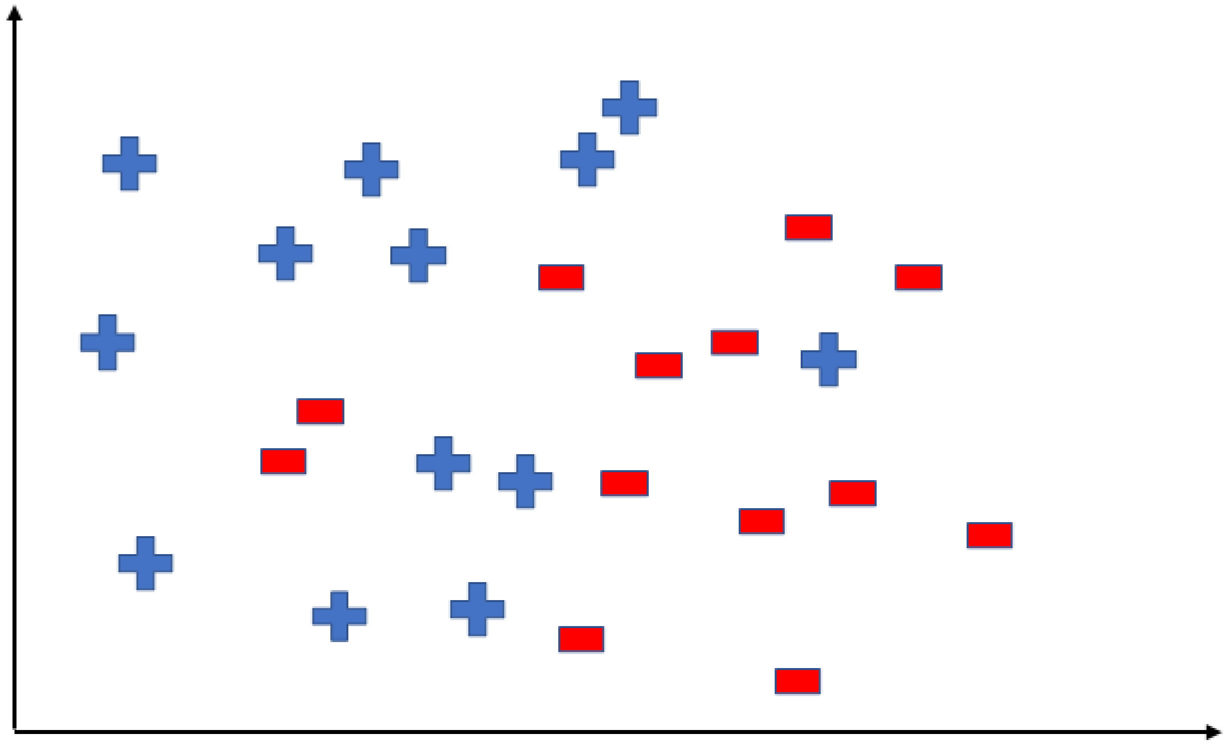
ML Understanding Hypothesis

In most supervised machine learning algorithm, our main goal is to find out a **possible hypothesis from the hypothesis space** that could possibly map out the inputs to the proper outputs.
The following figure shows the common method to find out the possible hypothesis from the Hypothesis space:

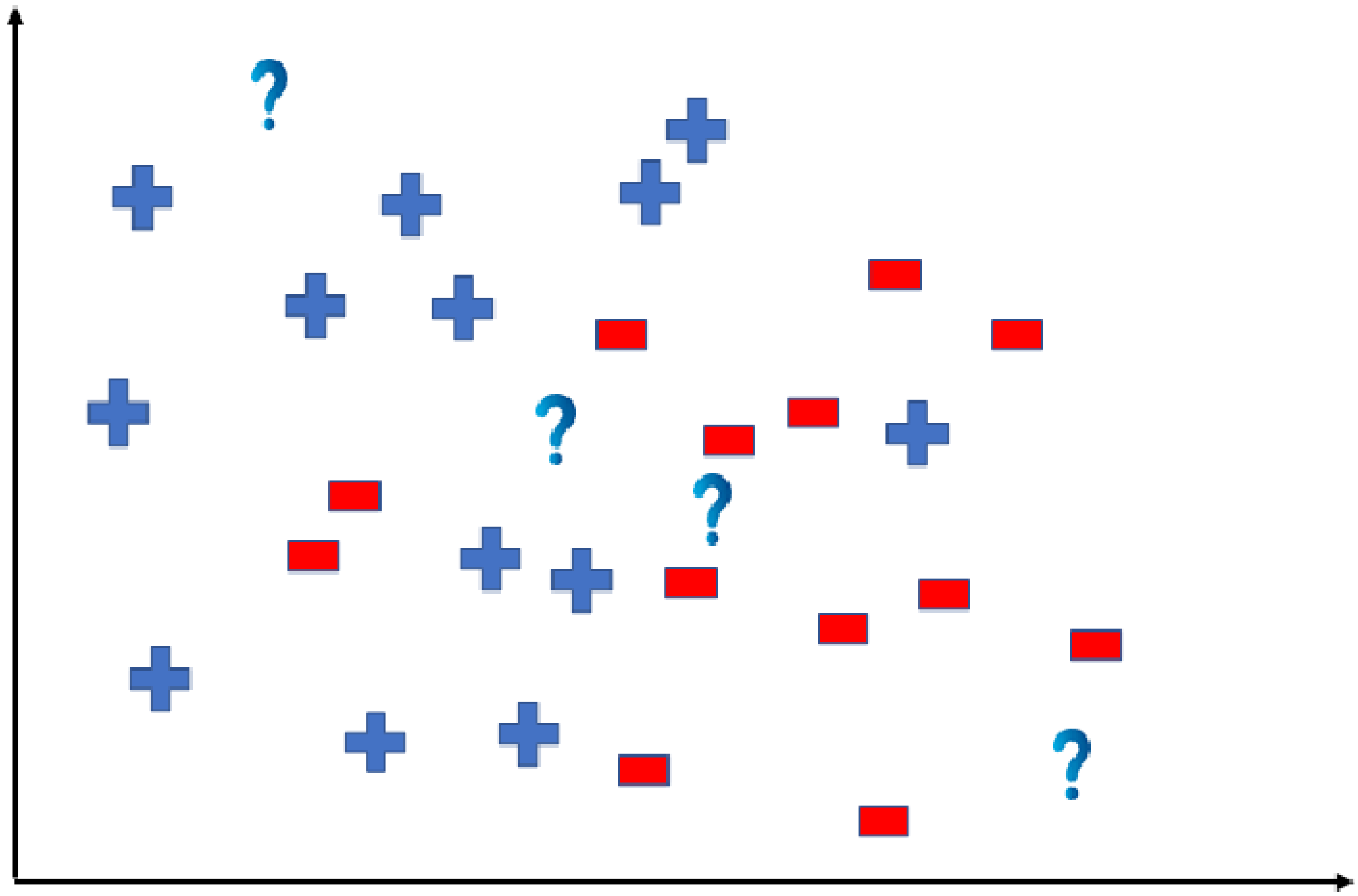


Hypothesis Space (H):
Hypothesis space is the **set of all the possible legal hypothesis**. This is the set from which the machine learning algorithm would determine the best possible (only one) which would best describe the target function or the outputs.
Hypothesis (h):
A hypothesis is a function that best describes the target in supervised

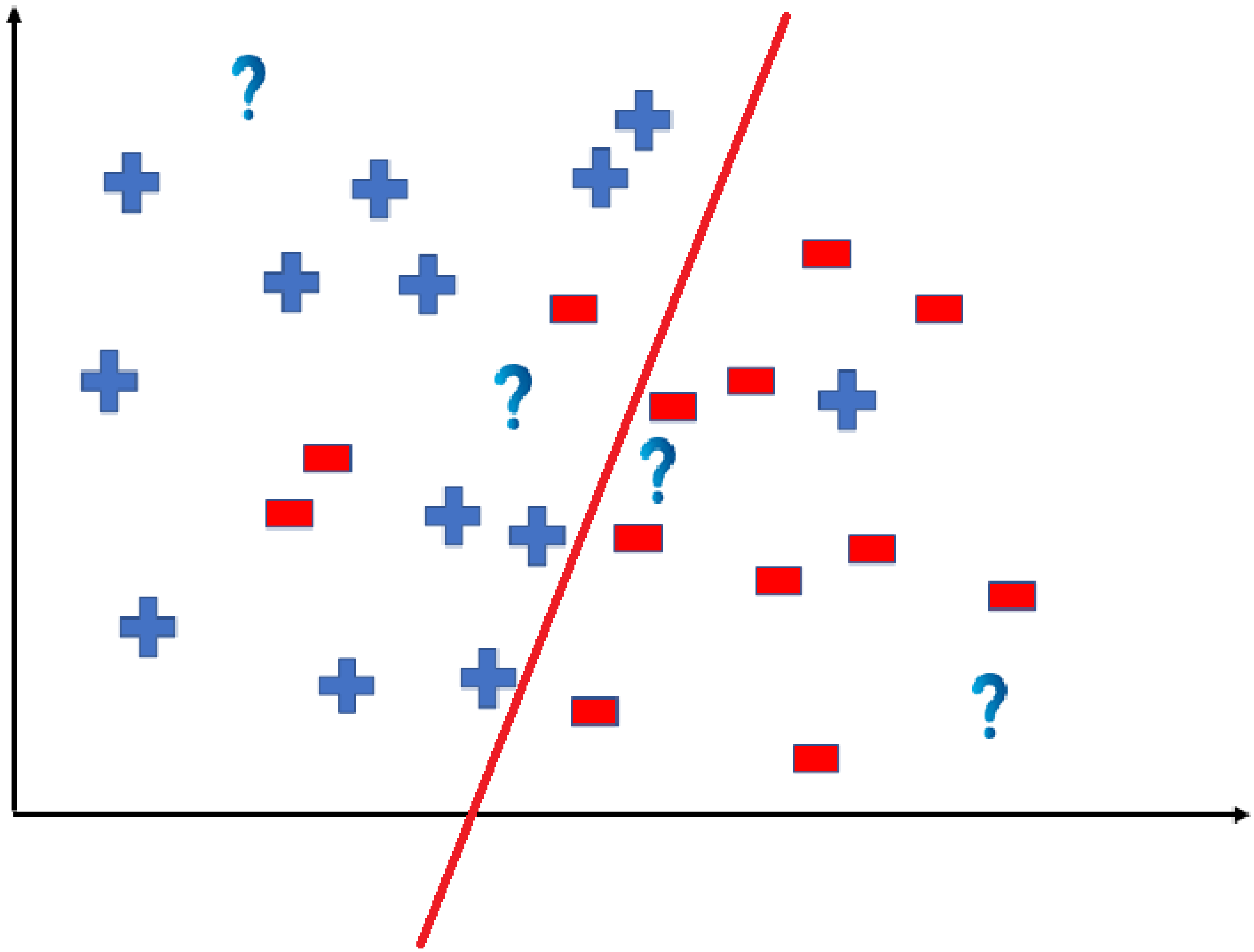
machine learning. The hypothesis that an algorithm would come up **depends upon the data and also depends upon the restrictions and bias** that we have imposed on the data. To better understand the Hypothesis Space and Hypothesis consider the following coordinate that shows the **distribution of some data:**



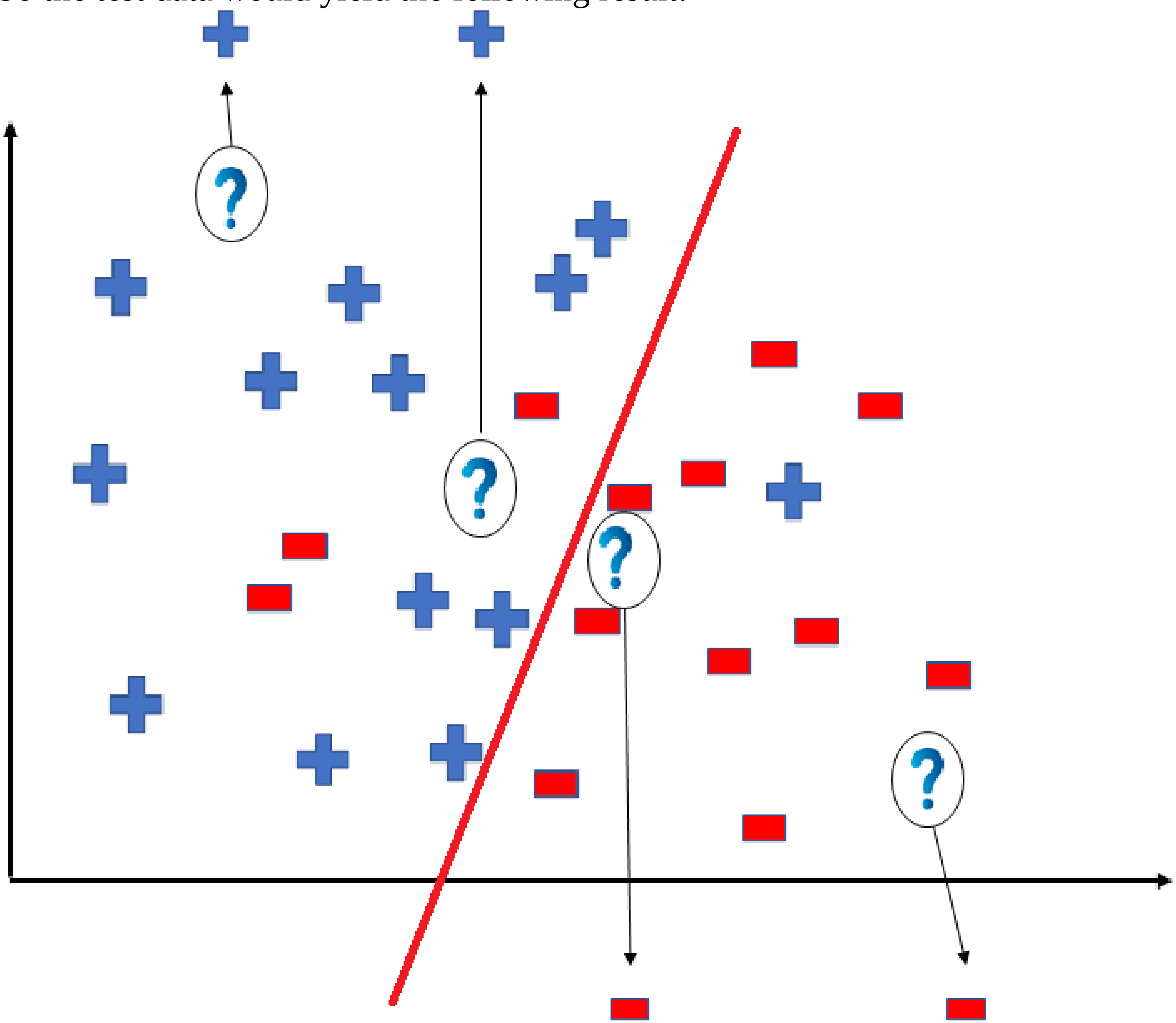
Say suppose we have test data for which we have to determine the outputs or results. The test data is as shown below:



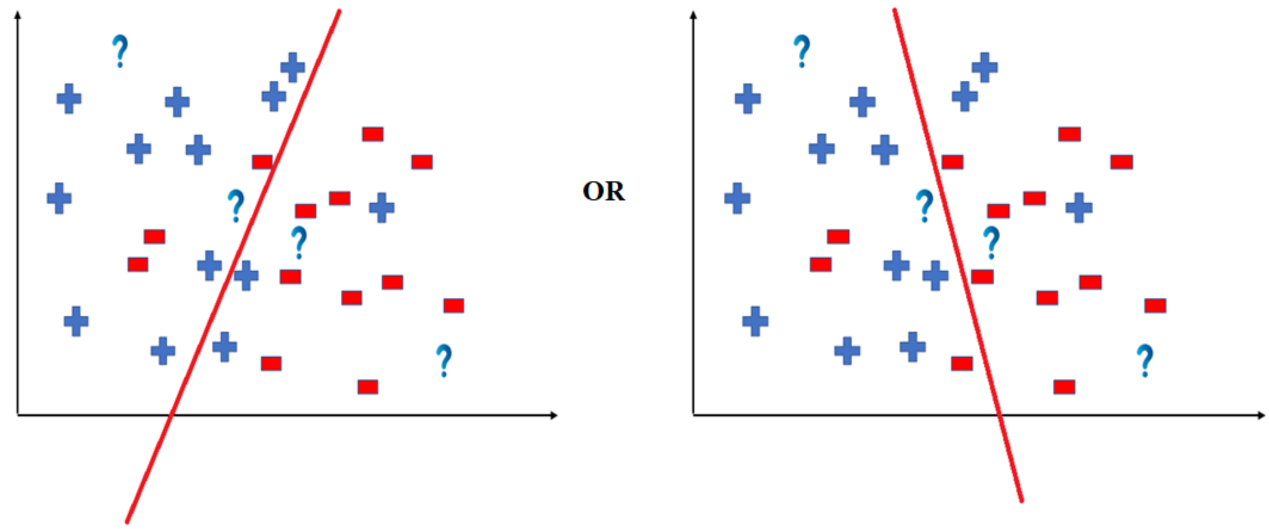
We can predict the outcomes by dividing the coordinate as shown below:



So the test data would yield the following result:

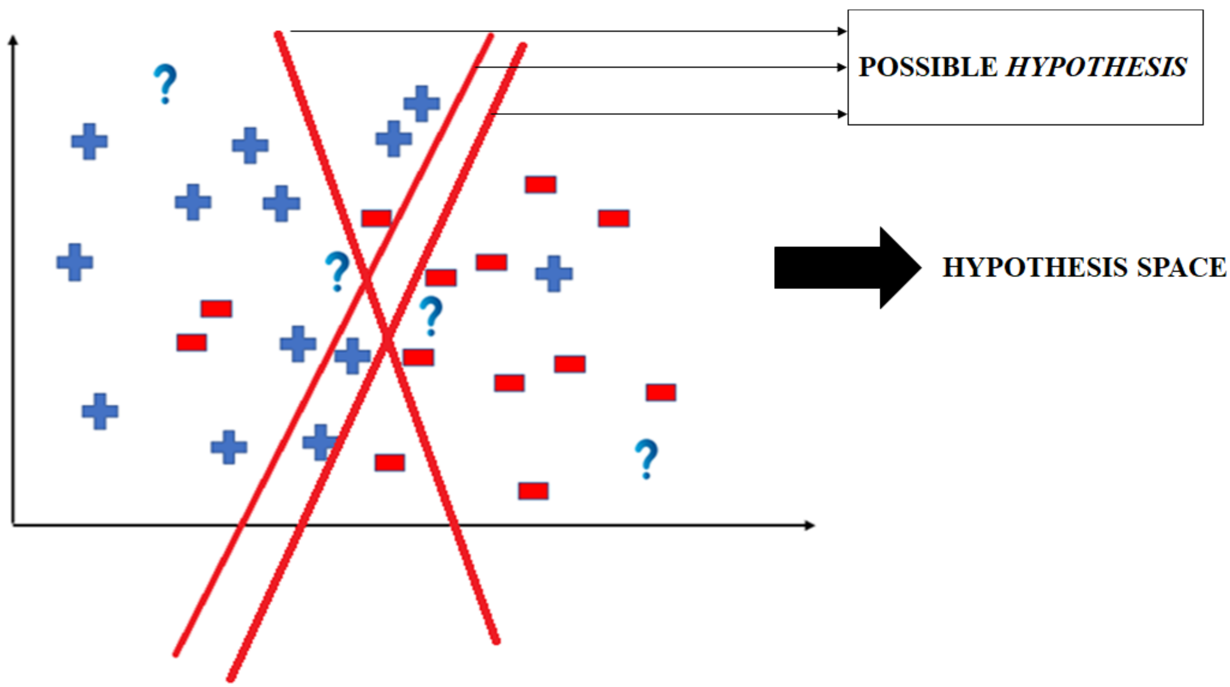


But note here that we could have divided the coordinate plane as:



The way in which the coordinate would be divided depends on the data, algorithm and constraints.

 All these legal possible ways in which we can **divide the coordinate plane to predict the outcome of the test data** composes of the Hypothesis Space.
 Each individual possible way is known as the hypothesis.
Hence, in this example the hypothesis space would be like:


POSSIBLE *HYPOTHESIS*
HYPOTHESIS SPACE

Hypothesis Testing in statistics

Hypothesis are statement about the given problem.

Hypothesis testing is a statistical method that is used in making a statistical decision using experimental data.

Hypothesis testing is basically an assumption that we make about a population parameter. It evaluates two mutually exclusive statements about a population to determine which statement is best supported by the sampledata

Example:

You say an **average student in the class is 30** or **a boy is taller than girls.** All those are an example in which we assume or need some statistic way to prove those. We need some mathematical conclusion whatever we are assuming is true.

Need for Hypothesis Testing

Hypothesis testing is an important procedure in statistics. Hypothesis testing evaluates two mutually exclusive population statements to determine which statement is most supported by sample data. When we say that the findings are statistically significant.

Parameters of hypothesis testing

- **Null hypothesis(H0):** In statistics, the null hypothesis is a general given statement or default position that there is no relationship between two measured cases or no relationship among groups. In other words, it is a basic assumption or made based on the problem knowledge.
- **This hypothesis is either rejected or not rejected based on the viability of the given population or sample**. Example: A company production is = 50 unit/per day etc.
- **Alternative hypothesis(H1):** The alternative hypothesis is the hypothesis used in hypothesis testing that is contrary to the null hypothesis.

Example : A company production is not equal to 50 unit/per day etc.

- **Level of significance**
  It refers to the degree of significance in which we accept or reject the null-hypothesis. 100% accuracy is not possible for accepting a hypothesis, so we, therefore, select a level of significance that is usually 5%. This is normally denoted with and generally, it is 0.05 or 5%, which means your output should be 95% confident(significance level is accepted) to give similar kind of result in each sample.
- **P-value**
  The P value, or calculated probability, is the probability of finding the observed/extreme results when the null hypothesis(H0) of a study given problem is true.
  -If your P-value is less than the chosen significance level then you reject the null hypothesis i.e. accept that your sample claims to support the alternative hypothesis.
  -P-**Value is a statistical measure, that helps to determine whether the hypothesis is correct or not.**

Example :

Given a coin and it is not known whether that is fair or tricky so let's decide null and alternate hypothesis

- Null Hypothesis(H0): a coin is a fair coin.
- Alternative Hypothesis(H1) : a coin is a tricky coin.

Now let's toss the coin and calculate p-value (probability value).

- Toss a coin 1st time and assume that result is head- P-value =50 (as head and tail have equal probability)
- Toss a coin 2nd time and assume that result again is head, now p-value = (1/2) * 50 ==50/2= 25

Error in Hypothesis Testing

- **Type I error:** When we reject the null hypothesis, although that hypothesis was true. Type I error is denoted by alpha.
- **Type II errors:** When we accept the null hypothesis but it is false. Type II errors are denoted by beta.

What is Generalization in Machine Learning?

Definition of generalization

In machine learning, generalization is a definition to demonstrate how well is a trained model to classify or forecast unseen data.

An **example** is when we train a model to **classify between dogs and cats.** If the model is provided with **dogs images dataset with only two breeds**, it may obtain a **good performance**.

But, it possibly gets a **low classification score** when it is tested by **other breeds of dogs as well.**

Therefore, data diversity(decrease redundancy) is very important factor in order to make a good prediction. In the sample above, the model may

obtain 85% performance score when it is tested by only two dog breeds and gains 70% if trained by all breeds.

However, the first possibly gets a very low score (e.g. 45%) if it is evaluated by an unseen dataset with all breed dogs. This for the latter can be unchanged given than it has been trained by high data diversity including all possible breeds.

It should be taken into account that data diversity is not the only point to care in order to have a generalized model.

In this post we explain all determinant factors. **There are some methods (regularization) to apply during model training to ensure about generalization. But before, we explain bias and variance as well as underfitting and overfitting.**

Variance and bias (overfitting and underfitting)

Variance and bias are two important terms in machine learning.

Variance means the **variety of predictions values** made by a machine learning model (target function). (Variance describes how much a random variable differs from its expected value.)

Bias means the distance of the predictions **from the actual (true) target** values (bias is the amount that a model's prediction differs from the target value, compared to the training data.)

A high-biased model means its **prediction values (average) are far from the actual values. (**High bias would cause an algorithm to miss relevant relations between the input features and the target outputs. This is sometimes referred to as underfitting.)

Also, high-variance **prediction means the prediction values are highly varied.(** a model that tries to fit most of the training dataset points making it complex)

Variance-bias trade-off

The prediction results of a machine learning model stand somewhere between

a) low-bias, low-variance,

b) low-bias, high-variance (overfit)

c) high-bias, low-variance, and

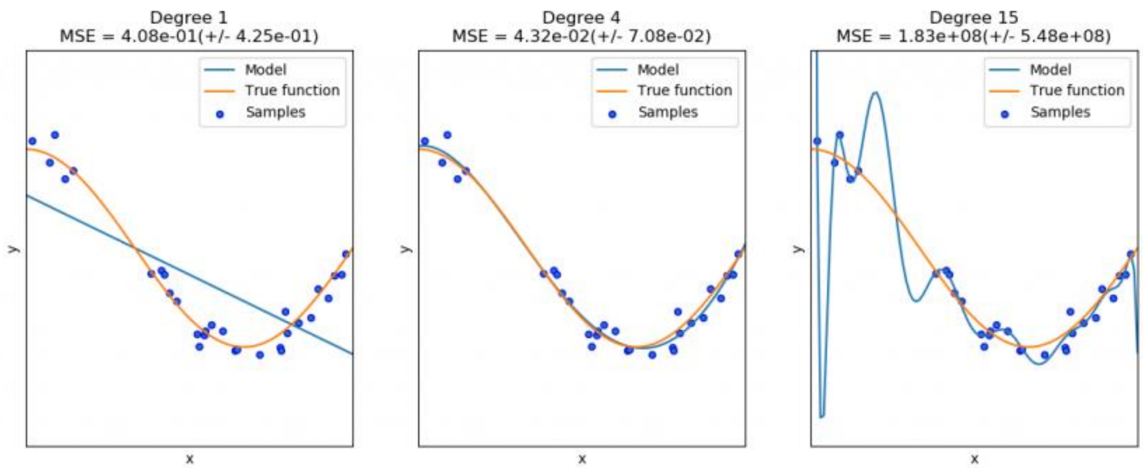
d) high-bias, high-variance.(underfit)

A low-biased, high-variance model is called **overfit** and a high-biased, low-variance model is called **underfit.**

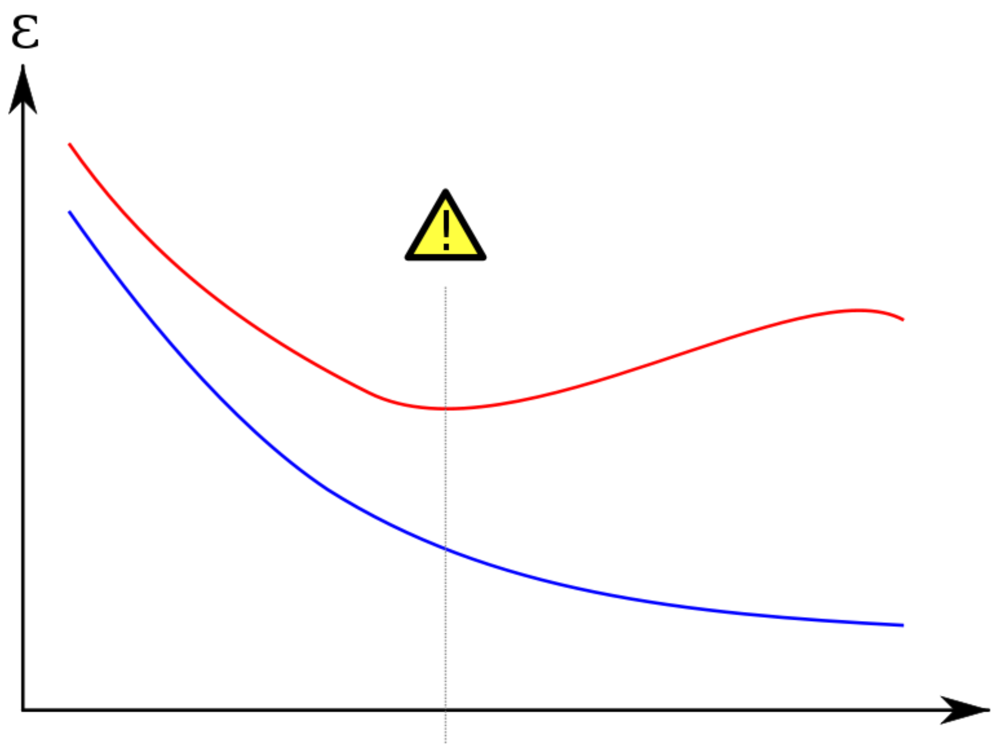
By generalization, we find the best trade-off between underfitting and overfitting so that a trained model obtains the best performance.

An overfit model obtains a high prediction score on seen data and low one from unseen datsets. (**data new to the model that was not part of the training**.)

An underfit model has low performance in both seen and unseen datasets.



Three models with underfitting (left), goodfit (middle), and overfitting (right). Credit: https://scikit-learn.org/



Overfitting/overtraining in supervised learning (e.g., neural network). **Training error is shown in blue, validation error in red, both as a function of the number of training cycles. I**f the validation error increases(positive slope) while the training error steadily decreases(negative slope) then a situation of overfitting may have occurred. The best predictive and fitted model would be where the validation error has its global minimum.

Determinant factors to train generalized models

There are different ways to secure that a machine learning model is generalized. Below we explain them.

Dataset

In order to train a classifier and generate a generalized machine learning model, a used dataset should contain diversity. It should be noted that it doesn't mean a huge dataset but a dataset containing all different samples. This helps classifier to be trained not only from a specific subset of data and therefore, the generalization is better fulfilled. In addition, during training, it is recommended to use

cross validation techniques such as K-fold or Monte-Carlo cross validations. These techniques better secure to exploit all possible portions of data and to **avoid generating an overfit model.**

Machine Learning algorithm

Machine learning algorithms differently act against overfitting, underfitting. **Overfitting** is more likely with nonlinear, non-parametric machine learning algorithms. For instance, **Decision Tree is a non-parametric machine learning algorithms, meaning its model is more likely with overfitting.** On the other hand, some machine learning models are too simple to capture complex underlying patterns in data. This cause to build an **underfit model. Examples are linear and logistic regression.**

Model complexity

When a machine learning models becomes too complex, it is usually prone to overfitting. There are **methods that help to make the model simpler.** They are called Regularization methods. Following we explain it.

Regularization

Regularization is collection of methods to make a machine learning model simpler. To this end, certain approaches are applied to different machine learning algorithms, for instance, **pruning for decision trees, dropout techniques for neural networks(reduction in overfitting), and adding a penalty parameters to the cost function in Regression.**

Bias and Variance in Machine Learning

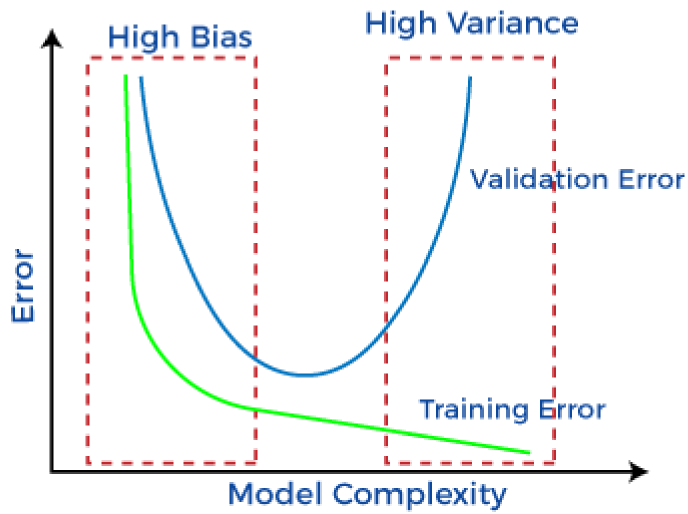
- Machine learning is a branch of Artificial Intelligence, which allows machines to Perform data analysis and make predictions.
- However, if the machine learning model is not accurate, it can make predictions errors,

and these **prediction errors** are usually known as **Bias and Variance**.

- In machine learning, these errors will always be present as there is always a slight difference between the **model predictions and actual predictions**.
- The main aim of ML/data science analysts is to reduce these errors in order to get more

accurate results. In this topic, we are going to discuss bias and variance, Bias-variance trade-off, Underfitting and Overfitting.

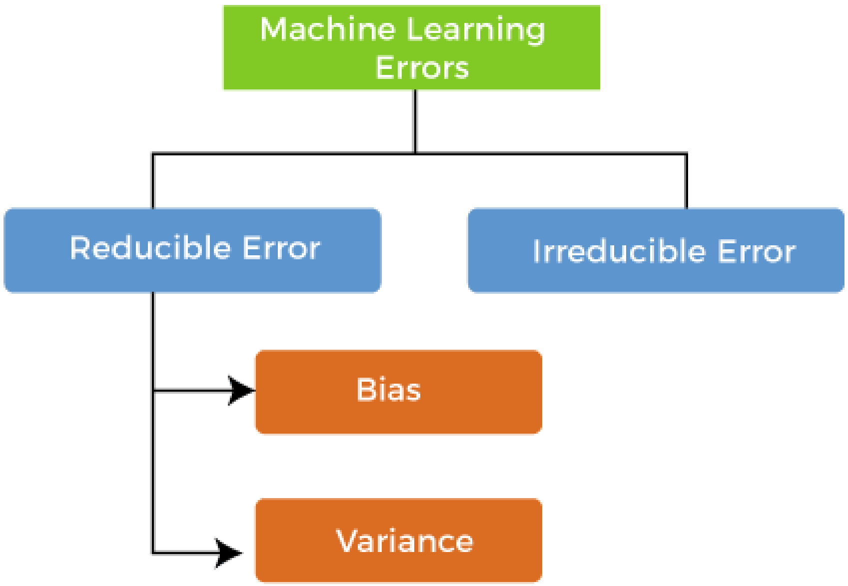
what errors in Machine learning are?



Errors in Machine Learning?

In machine learning, an error is a measure of how accurately an algorithm can make predictions for the previously unknown dataset. On the basis of these errors, the machine learning model is selected that can perform best on the particular dataset. There are mainly **two types of errors in machine learning**, which are:

- **Reducible errors:** These errors can be reduced to improve the model accuracy. Such errors can further be classified into bias and Variance.
- **Irreducible errors:** These errors will always be present in the model regardless of which algorithm has been used. The cause of these errors is unknown variables whose value can't be reduced.
-



What is Bias?

In general, a machine learning model analyses the data, find patterns in it and make predictions.

While training, the model learns these patterns in the dataset and applies them to test data for prediction.

While making predictions, a difference occurs between prediction values made by the model and actual values/expected values**, **and this difference is known as bias errors or Errors due to bias.

It can be defined as an inability of machine learning algorithms such as Linear Regression to capture the true relationship between the data points.

Each algorithm begins with some amount of bias because bias occurs from assumptions in the model, which makes the target function simple to learn. A model has either:

- **Low Bias:** A low bias model will make fewer assumptions about the form of the target function.
- **High Bias:** A model with a high bias makes more assumptions, and the model becomes unable to capture the important features of our dataset. **A high bias model also cannot perform well on new data.**

Generally, a linear algorithm has a high bias, as it makes them learn fast. The simpler the algorithm, the higher the bias it has likely to be introduced. Whereas a nonlinear algorithm often has low bias.

Some examples of machine learning algorithms with low bias **are Decision Trees, k-Nearest Neighbours and Support Vector Machines**. At the same time, an algorithm with high bias is **Linear Regression, Linear Discriminant Analysis and Logistic Regression.**

Ways to reduce High Bias:

High bias mainly occurs due to a much simple model. Below are some ways to reduce the high bias:

- Increase the input features as the model is underfitted.
- Decrease the regularization term.
- Use more complex models, such as including some polynomial features.

What is a Variance Error?

The variance would specify the amount of variation in the prediction if the different training data was used. In simple words, ***variance tells that how much a random variable is different from its expected value.*** Ideally, a model should not vary too much from one training dataset to another, which means the algorithm should be good in understanding the hidden mapping between inputs and output variables. Variance errors are either of **low variance or high variance.**

Low variance means there is a small variation in the prediction of the target function with changes in the training data set.

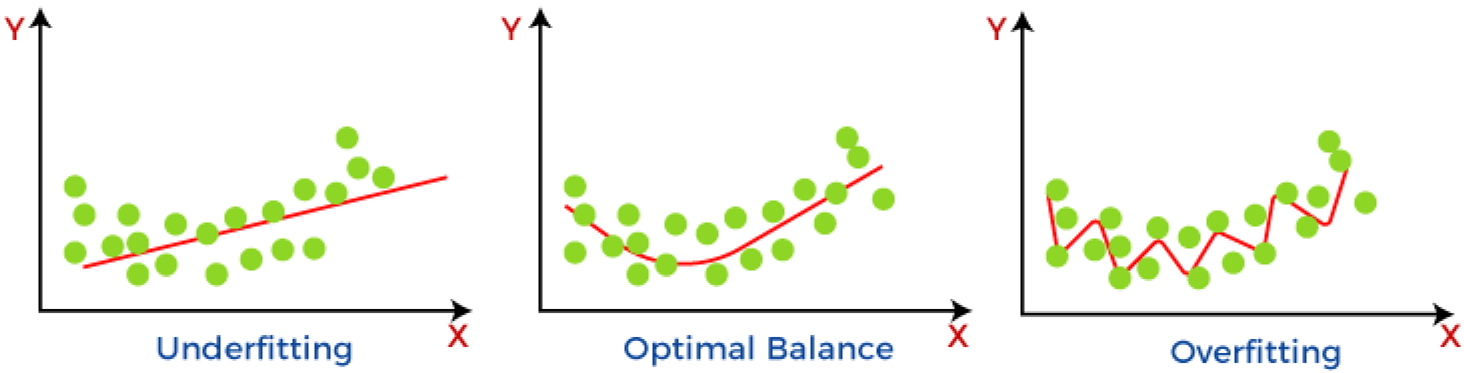
At the same time, **High variance** shows a large variation in the prediction of the target function with changes in the training dataset.

A model that shows high variance learns a lot and perform well with the training dataset, and does not generalize well with the unseen dataset. As a result, such a model gives good results with the training dataset but shows high error rates on the test dataset.

Since, with high variance, the model learns too much from the dataset, it leads to overfitting of the model. **A model with high variance has the below problems:**

- A high variance model leads to overfitting.
- Increase model complexities.

Usually, nonlinear algorithms have a lot of flexibility to fit the model, have high variance.



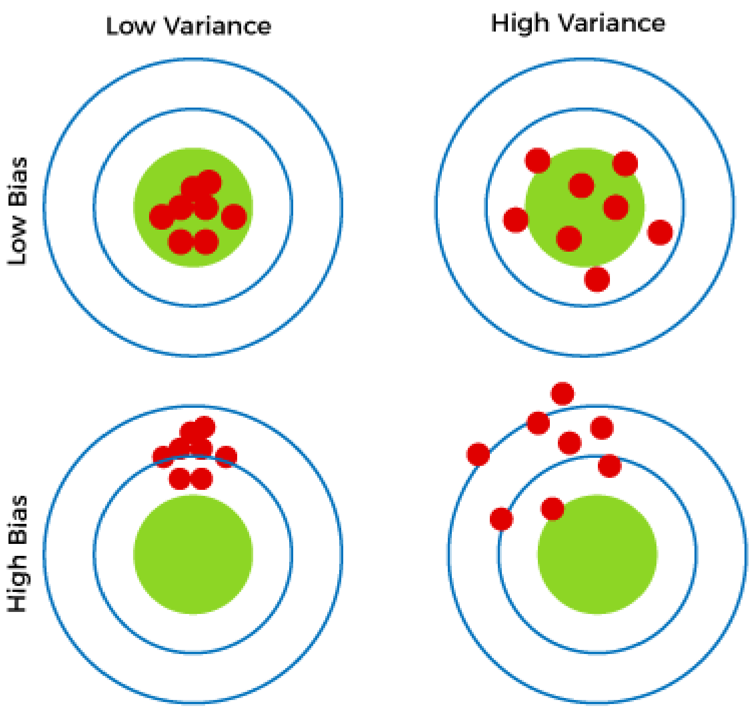
Some examples of machine learning algorithms with low variance are, **Linear Regression, Logistic Regression, and Linear discriminant analysis**. At the same time, algorithms with hi variance are **decision tree, Support Vector Machine, and K-nearest neighbours.**

Ways to Reduce High Variance:

- Reduce the input features or number of parameters as a model is overfitted.
- Do not use a much complex model.
- Increase the training data.
- Increase the Regularization term.

Different Combinations of Bias-Variance

There are four possible combinations of bias and variances, which are represented by the below diagram:

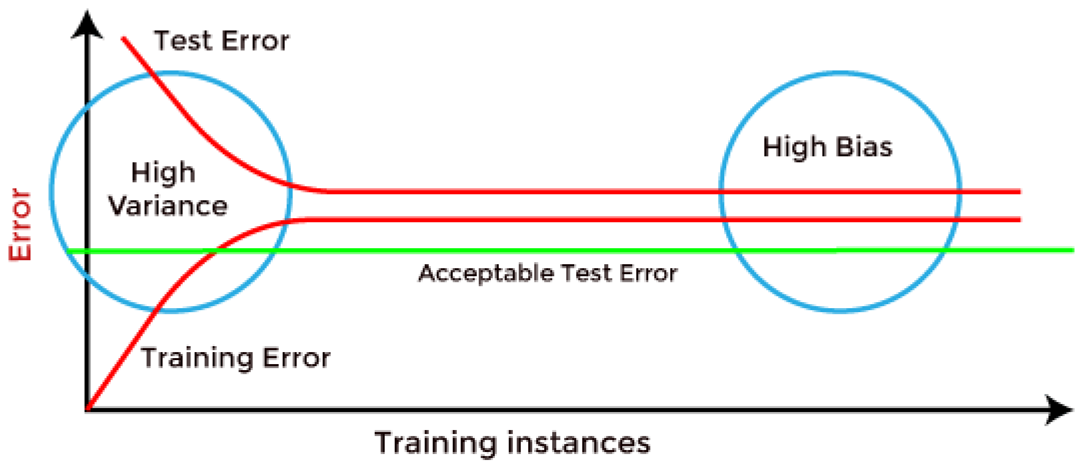


1. **Low-Bias,LowVariance**
   The combination of low bias and low variance shows an ideal machine learning model. However, it is not possible practically.
2. **Low-Bias, High-Variance:** With low bias and high variance, model predictions are inconsistent and accurate on average. This case occurs when the model learns with a large number of parameters and hence leads to an **overfitting**
3. **High-Bias, Low-Variance:** With High bias and low variance, predictions are consistent but inaccurate on average. This case occurs when a model does not learn well with the training dataset or uses few numbers of the parameter. It leads to **underfitting** problems in the model.
4. **HighBias,HighVariance**

With high bias and high variance, predictions are inconsistent and also inaccurate on average.

How to identify High variance or High Bias?

High variance can be identified if the model has:



- Low training error and high test error.

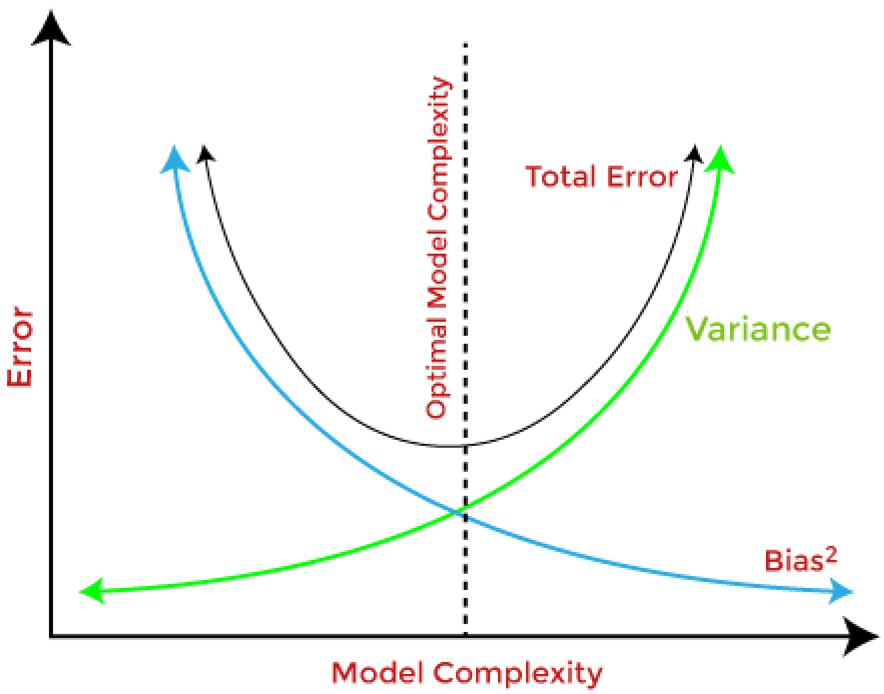
High Bias can be identified if the model has:

- High training error and the test error is almost similar to training error.

Bias-Variance Trade-Off

While building the machine learning model, it is really important to take care of bias and variance in order to avoid overfitting and underfitting in the model. If the model is very simple with fewer parameters, it may have low variance and high bias. Whereas, if the model has a large number of parameters, it will have high variance and low bias.

So, **it is required to make a balance between bias and variance errors, and this balance between the bias error and variance error is known as the Bias-Variance trade-off.**



For an accurate prediction of the model, algorithms need a low variance and low bias.

But this is not possible because bias and variance are related to each other:

- If we decrease the variance, it will increase the bias.
- If we decrease the bias, it will increase the variance.

Bias-Variance trade-off is a central issue in supervised learning. Ideally, we need a model that accurately captures the regularities in training data and simultaneously generalizes well with the unseen dataset. Unfortunately, doing this is not possible simultaneously. Because a high variance algorithm may perform well with training data, but it may lead to overfitting to noisy data. Whereas, high bias algorithm generates a much simple model that may not even capture important regularities in the data. So, we need to find a sweet spot between bias and variance to make an optimal model.

Hence, the ***Bias-Variance trade-off is about finding the sweet spot to make a balance between bias and variance errors.***