

# Ch $\Rightarrow$ 1. Introduction:

## An Overview of Statistical Learning

Statistical learning refers to a vast set of tools for understanding data.

It is a framework for Machine Learning drawn from the field of statistics & functional Analysis.

Statistical Learning Theory deals with the problem of finding a predictive function based on data.

These tools can be classified as:

- ① Supervised Learning: involves tasks like prediction (classification or regression) on an output based learning technique
- ② Unsupervised learning: involves tasks like ~~distorting~~ clustering where no supervising is done i.e. no output is present.

## Applications:

1. Identify the risk factors for prostate cancer.

This task includes finding the parameters that cause or lead to prostate cancer.



(2) Classifying the sound of 2 vowels.

we can get a graph of Frequency v/s periodogram  
and to classify the 2 classes.  
(This task includes good need of pre-processing  
of data)

(3) Prediction of Heart Attack

we can do this using parameters like food,  
geographic location, tobacco etc.

(4) Email Spam Detection

Where every word is actually treated as  
parameters & classify them.

(5) Gene Expression Data.

Checking what <sup>gene</sup> data is active at a time  
when cancer is there & predicting  
cancer on that basis.

Etc.



## Supervised Learning :

1. Vectors of " $X$ " predictors of  $X$  to predict output " $Y$ " also called 'target'.
2. In regression problem,  $Y$  is quantitative (price, BPX)
3. In classification problem,  $Y$  is qualitative (Malignant, Benign, dog/cat etc)

## Responsibility :

4. Understanding the inputs that affect outputs
5. Test our model on unseen cases
6. Assess the quality of our predictions & inferences

This is called supervised learning because:

the technique is supervised by outputs

Example: a child is taught shapes  
multiple images of squares & triangles

& the child was tested by giving a ~~bag~~  
such shape him some toys &  
classify its shape.



on the other hand, if the same kid is shown any images or is given to classify those, shapes when a bag full of such toys are given.

The child wouldn't know if the shape is triangle or square but could classify that there are 2 different shapes.

## Unsupervised Learning:

1. No "Y" is given.
2. Objective is to find similar type of data, find features that behaves similarly, find linear combination of features with most variation.

## Challenges:

3. Difficult to analyze the situation.
4. Can be used in semi-supervised learning.



# Statistical Learning V/s Machine Learning

Machine Learning is a subfield of AI.

Statistical Learning V/s is a subfield of Stats.

They both have a lot in common.

1. Both works on Supervised & Unsupervised Learning.

ML has emphasis on large scale application & prediction accuracy.

SL emphasizes on V/s Models & their precision & uncertainty.