

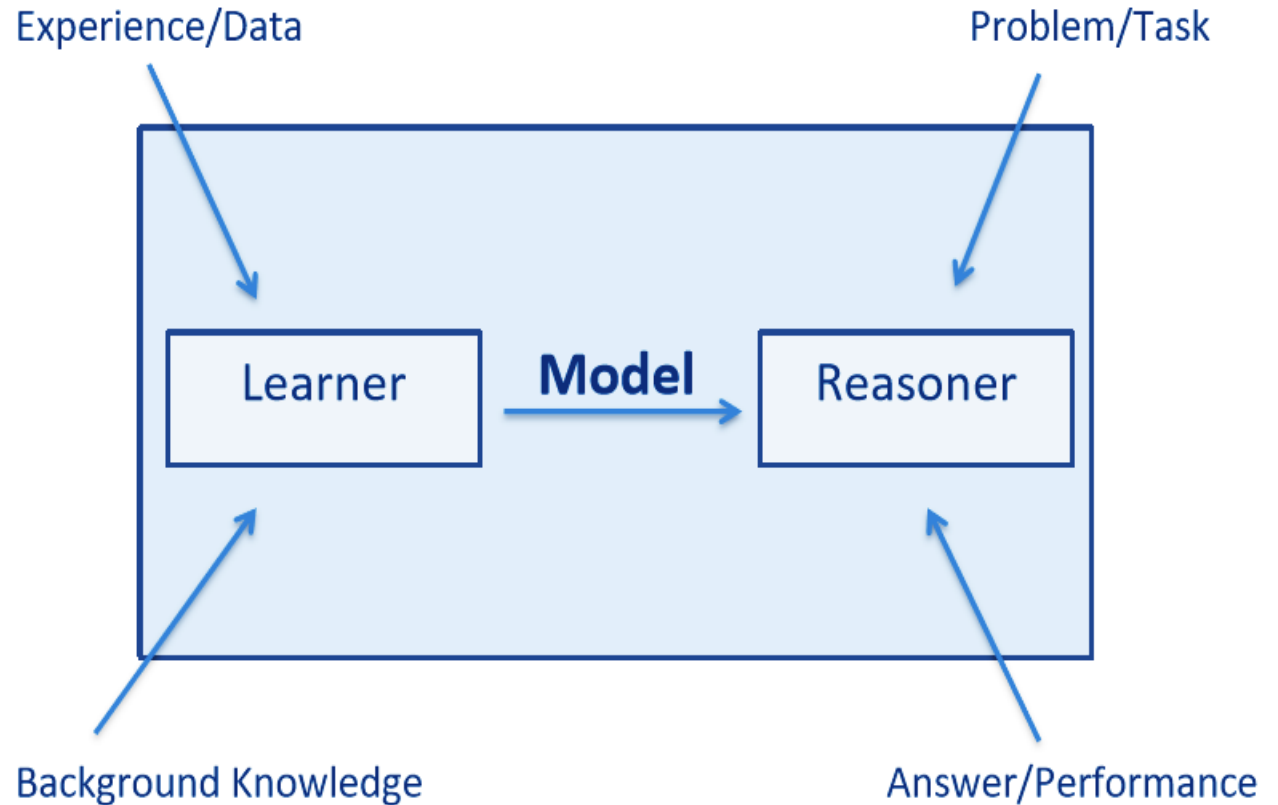
# MACHINE LEARNING

# Learning

- ◇ Learning is the ability to improve behavior based on experience
- ◇ Machine Learning:
  - It will explores algorithms that, build / learn models for data
  - Model used for prediction, cluster making or solving tasks

- ◇ T:
  - Prediction, Classification (Supervised)
  - Cluster Generation (Unsupervised)
  - Acting in an environment (Reinforcement)
  
- ◇ E: data
  
- ◇ Performance Measure , P:
  - Increasing accuracy in problem solving
  - Adding new skills to the agent
  - Increasing reward
  - Improves efficiency of problem solving

# ML System: ML system is like a box



# ML System: ML system is like a box

- ◇ It feed with a data or experience, there is a problems or tasks that requires solution. In addition we can also give background knowledge which helps the system.
- ◇ For the problem or task the learning program comes up with a solution or answer in such a way so that the corresponding performance can be measured.
- ◇ Inside the system there are two components
  - The learner
  - reasoner
- ◇ The learner tasks the data or experience, he can also consider background knowledge and from this the learner builds model.
- ◇ This model can be used by the reasoner who is given a task, needs to find a solution for that task
  
- ◇ So, the learner tasks the background knowledgw and experience and learns.

# How to create a learner:

To create a learner the following steps are required:

1. Choose the training experience
2. Choose the target function to represent the model
3. Choose how to represent the target function
4. Finally, choose a learning algorithm to inter the target function

# Feature and their types

Often individual observation are analyzed into a set of quantifiable properties which are called features.

1. Categorical
2. Ordinal
3. Integer value
4. Real value

## Terminology

- **Example  $(x,y)$ :** Instance  $x$  with label  $y=f(x)$ .
  - **Training Data  $S$ :** Collection of examples observed by learning algorithm.
  - **Instance Space  $X$ :** Set of all possible objects describable by features.
  - **Concept  $c$ :** Subset of objects from  $X$  ( $c$  is unknown).
  - **Target Function  $f$ :** Maps each instance  $x \in X$  to target label  $y \in Y$
-

# Hypothesis space:

- ◇ One way to think about a supervised learning machine is as a device that explores a hypothesis space.
- ◇ Give the type of function the hypothesis space is the set of candidate outputs i.e. suppose to be obtained using a training function.
- ◇ Each settings of the parameters in the machine is the different hypothesis about the function that maps input vectors to output vectors.

# Some terminologies for hypothesis space

- Hypothesis  $h$ : Function that approximates  $f$ .
- Hypothesis Space  $\mathcal{H}$  : Set of functions we allow for approximating  $f$ .
- The set of hypotheses that can be produced, can be restricted further by specifying a language bias.
- Input: Training set  $\mathcal{S} \subseteq \mathcal{X}$
- Output: A hypothesis  $h \in \mathcal{H}$

# Representation:

1. Decision tree
2. Linear Function
3. Multivariate Linear Function
4. Single Layer perceptron
5. Multilayer neural network

# Inductive Bias

Need to make assumption

- Experience alone doesn't allow us to make conclusions about unseen data instances.

Two types of bias:

- Restriction: limit the hypothesis space.
- Preference: Impose ordering on hypothesis space.

# Inductive learning

Inducing a general function from training example.

- Construct hypothesis  $h$  to agree with  $c$  on the training examples.
- A hypothesis is consistent if it agrees with all training examples.
- A hypothesis is said to be generalized well if it correctly predict the value of  $y$  for novel examples.

# Inductive learning hypothesis

Any hypothesis  $h$  found to approximate the target function  $C$  well over a sufficiently large set of training example  $D$  will also approximate the target function well over other unobserved example.

## Generalization:

The components of generalization error:

- Bias : How much the average model over all training sets differ from the true model ?
- Variance : How much models estimated from different training sets differ from each other.

# Underfitting and Overfitting

## Underfitting :

Model is too simple represent all the relevant class characteristics.

- High bias and low variance.
- High training error and high testing error.

## Overfitting :

Model is too complex and fits irrelevant characteristics(noise) in the data.

- Low bias and high variance.
- Low training error and high testing error.

# Experimental evaluation of learning algorithm

*Evaluating the performance of learning algorithm is important because learning systems are usually designed to predict the class of “future” unlabeled data points*

- Typical choices for performance evaluation..
  - Error
  - Accuracy
  - Precision/Recall
- Typical choices for sampling method :
  - Train or test set
  - k-fold cross validation

# Error Measurements:

## Sample Error and True Error

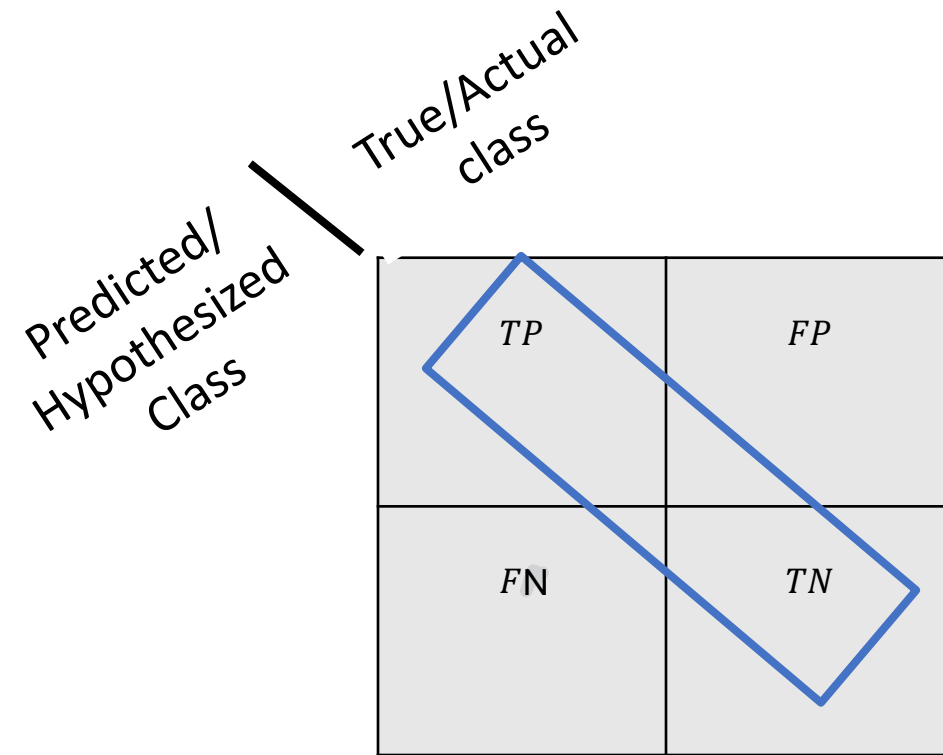
- The **sample error** of hypothesis  $f$  with respect to target function  $c$  and data sample  $S$  is:

$$\text{error}_S(f) = 1/n \sum_{x \in S} \delta(f(x), c(x)) \quad \text{where, } \delta(f(x), c(x)) = \begin{cases} 1 & \text{when } f(x) \neq c(x) \\ 0 & \text{when } f(x) = c(x) \end{cases}$$

- The **true error** (denoted  $\text{error}_D(f)$ ) of hypothesis  $f$  with respect to target function  $c$  and distribution  $D$ , is the probability that  $h$  will misclassify an instance drawn at random according to  $D$ .

$$\text{error}_D(f) = \Pr_{x \in D} [f(x) \neq c(x)]$$

# Confusion Matrix



# Confusion Matrix

## True Position ( TP ) :

TP= those instances for which the true class is positive and user also hypothesized or predicted positive, they can be called true positive.

## True Negative ( TN ) :

TN= those instances where the true negative classes are also output as negative by users prediction

TP & TN are the zones where the learning algorithm predicts correctly.

## False Positive :

This means the instances are actually negative but the learning algorithm is wrongly classifying them as positive .

## False Negative :

This means the learning algorithm erroneously marks them as negative but those instances should have been positive.

$$\text{Accuracy} = \frac{TP+TN}{P+N}$$

## Precision :

Precision is define to be out of the examples that the learning algorithm marks as positive, how many are correctly positive.

$$\text{Precision} = \frac{TP}{TP+FP}$$

**Recall** : (True positive rate)

This measures how many of the positive examples the learning algorithm retrieves as positive.

$$\text{Recall} = \frac{TP}{TP+FN}$$

### **Homework**

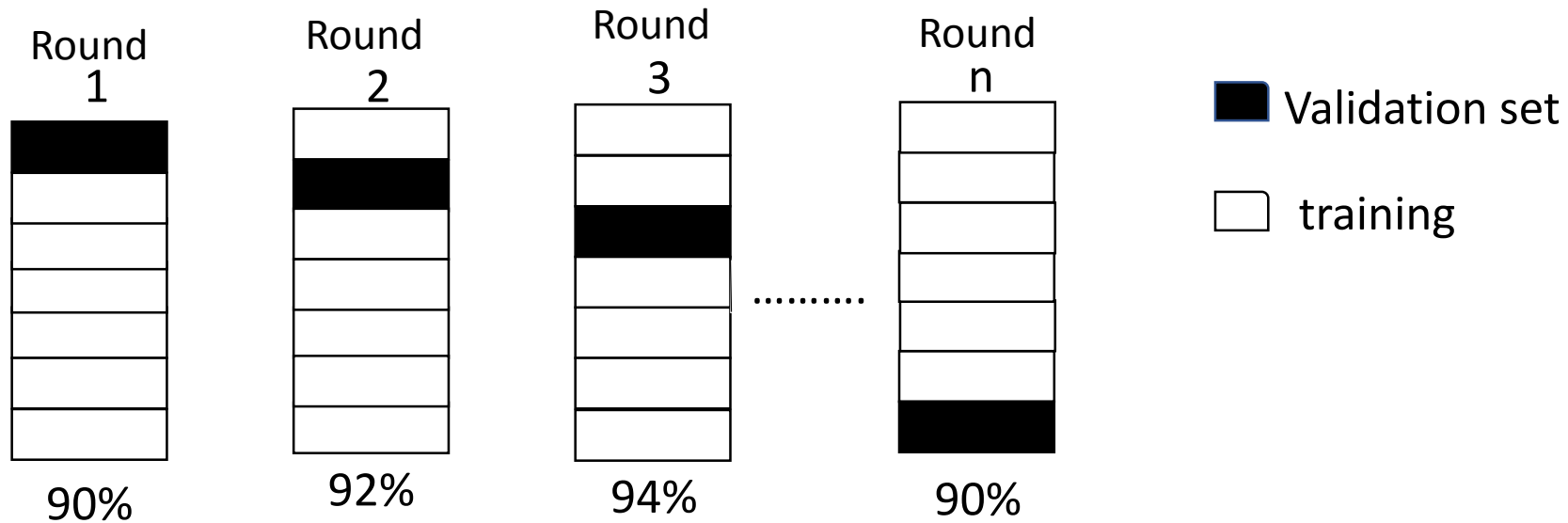
- # Supervised learning Vs Unsupervised learning
- # Types of attributes & their differences
- # Classification Vs Regression
- # Bias Vs Variance
- # Overfitting Vs Underfitting
- # Continuous attributes Vs categorical attributes

# Difference between Bias and Variance

<b>Subject</b>	<b>Bias</b>	<b>Variance</b>
1.Definaton	Erroneous assumption in the learning algorithm.	Sensitivity of model towards noise rather than important feature relationship between input and output
2.Increase in the number of features	Decreases	Increases
3.Increase in the number of parameters	Decreases	Increases
4.Increase in training examples	Remains the same training examples	Decreases

# Steps of k-fold cross validation

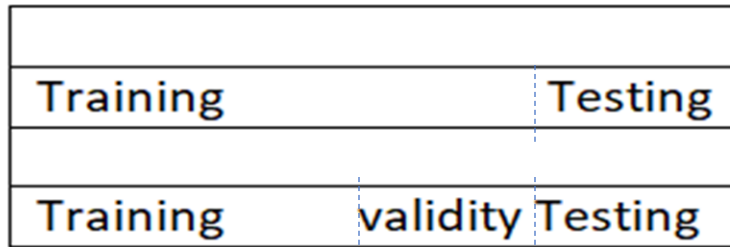
- Split the data into k equal subset
- Perform k rounds of learning on each round-
  - a.  $1/k$  of the data is consider as test set
  - b. The remaining examples are used as training set
- Compute the average test set score of the k rounds



Final Accuracy = Average accuracy for all rounds

# Validation set inside a training set

Original Set



Train the Machine

Evaluate generalization/

Do final evaluation

Used for **tuning** the model

A large, thick black L-shaped frame surrounds the central text. The top-left corner is a horizontal bar extending to the right, then a vertical bar extending downwards. The bottom-right corner is a horizontal bar extending to the left, then a vertical bar extending upwards.

# CONFUSION MATRIX

Data Mining

# Definition

- A confusion matrix is a table that is often used to describe the performance of a classification model (or “classifier”) on a set of test data for which the true values are known. It allows the visualization of the performance of an algorithm.

# Definition of the Terms

- Positive (P) : Observation is positive (**for example: is an apple**).
- Negative (N) : Observation is not positive (**for example: is not an apple**).
- True Positive (TP) : Observation is positive, and is predicted to be positive.
- False Negative (FN) : Observation is positive, but is predicted negative.
- True Negative (TN) : Observation is negative, and is predicted to be negative.
- False Positive (FP) : Observation is negative, but is predicted positive.

# 1. Classification Rate/Accuracy

- However, there are problems with accuracy. It assumes equal costs for both kinds of errors.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- A 99% accuracy can be excellent, good, mediocre, poor or terrible depending upon the problem.

# 2. Recall

- Recall can be defined as the ratio of the total number of correctly classified positive examples divide to the total number of positive examples. High Recall indicates the class is correctly recognized (small number of FN).

$$\mathbf{Recall} = \frac{TP}{TP + FN}$$

# 3.Precision

To get the value of precision we divide the total number of correctly classified positive examples by the total number of predicted positive examples. High Precision indicates an example labeled as positive is indeed positive (small number of FP).

Precision is given by the relation:

$$\mathbf{Precision} = \frac{TP}{TP + FP}$$

# Process

## How to Calculate a Confusion Matrix

- ✓ Step 1) First, you need to test dataset with its expected outcome values.
- ✓ Step 2) Predict all the rows in the test dataset.
- ✓ Step 3) Calculate the expected predictions and outcomes:

# Examples: Confusion Matrix

A dataset has **125 data** from positive(+) class and **75 data** from negative class. For this dataset,

Model-1 gives **150 + class** and **50 - class**.

And

Model-2 gives **120 + class** and **80 - class**.

Find accuracy, Precision and recall of these models.

A dataset has **125 data** from positive(+) class and **75 data** from negative class. For this dataset, Model-1 gives 150 + class and 50 - class. And Model-2 gives 120 + class and 80 - class.

We Know That,

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

For model 1

In this dataset ,

T.P (True Positives) is =**125**

T.N (True Negatives) is **75**

F.P (False Positives) =(150-125)=25

and

F.N = (75-50)=25

- So precision=**125**/(125+25)=.83 (Ans)
- Recall=**125**/(125+25)=.83 (Ans)
- Accuracy=**125+75**/(125+75+25+25)=.8 (Ans)

A dataset has 125 data from positive(+) class and 75 data from negative class. For this dataset, Model-1 gives 150 + class and 50 - class. And Model-2 gives 120 + class and 80 - class.

We Know That,

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

In this dataset ,

T.P (True Positives) is =125

T.N (True Negatives) is 75

For model 2

$$\text{F.P} = (125-120)=5 \text{ and } \text{F.N} = (80-75)=5$$

- Precision=  $125/(125+5)=.96$  (Ans)
- Recall=  $.96$  (Ans)
- Accuracy =  $(125+75)/(125+75+5+5)=.952$  (Ans)

# 3X3 Confusion Matrix Examples

**Calculate Precision , Recall per Class and also Calculate Accuracy**

Actual	Predicted		
	15	2	3
	7	15	8
	2	3	45

		Predicted			Total
		A	B	C	
Actual	A	15	2	3	20
	B	7	15	8	30
	C	2	3	45	50
Total		24	20	56	100

➤ Precision  
= Correctly Predicted / Total Predicted

- ✓ Class A Precision =  $15 / 24 = 0.62$
- ✓ Class B Precision =  $15 / 20 = 0.75$
- ✓ Class C Precision =  $45 / 57 = 0.80$

Actual	Predicted			Total
	A	B	C	
A	15	2	3	20
B	7	15	8	30
C	2	3	45	50
Total	24	20	56	100

➤ Recall  
= Correctly Classified / Total Actual

$$\text{Class A Recall} = 15 / 20 = 0.75$$

$$\text{Class B Recall} = 15 / 30 = 0.5$$

$$\text{Class C Recall} = 45 / 50 = 0.9$$

Actual	Predicted			Total
	A	B	C	
A	15	2	3	20
B	7	15	8	30
C	2	3	45	50
Total	24	20	56	100

➤ Accuracy  
= Total Correctly Classified / Total  
(Actual and Predicted)

✓ Accuracy of Classified =  $(15 + 15 + 45) / 100$   
= 0.75

# 5X5 Confusion Matrix

## 4X4 Confusion Matrix

are also Same Method to Calculate Precision , Recall per Class and also Calculate Accuracy

✓ Precision

= Correctly Predicted / Total Predicted

✓ Recall

= Correctly Classified / Total Actual

✓ Accuracy

= Total Correctly Classified / Total (Actual and Predicted)

# Why We Need Confusion Matrix?

- It shows how any classification model is confused when it makes predictions.
- This breakdown helps you to overcome the limitation of using classification accuracy alone.
- Every column of the confusion matrix represents the instances of that predicted class.
- Each row of the confusion matrix represents the instances of the actual class.

# Advantages

- Accommodates a wide variety of project-oriented business activities.
- Provides good training ground for strategic managers
- Maximizes efficient use of functional managers.

# Disadvantages

- Causes participant to experience dual authority, which can be frustrating and confusing.
- Means participant need good interpersonal skills and extensive training.
- May result in confusion and contradictory policies.