PH-227
AI and Data Science

- Aftab Alam
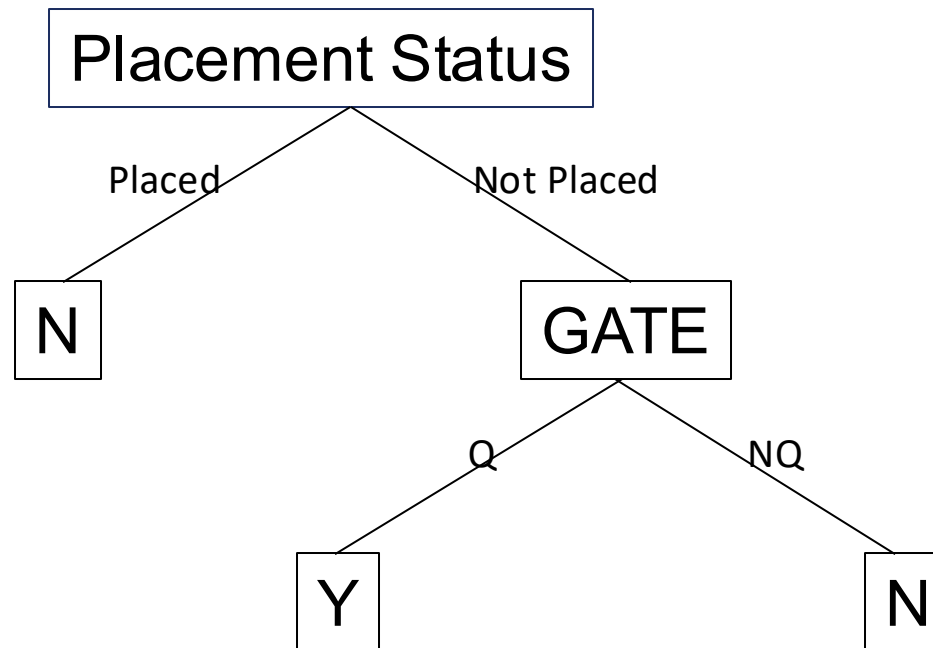- Email : aftab@iitb.ac.in
- Ext. : 5564 or 8564

TAs:

Yashowardhan, Divyansh, Matam, Peela, Piyush

A Quick Recap

Decision Tree

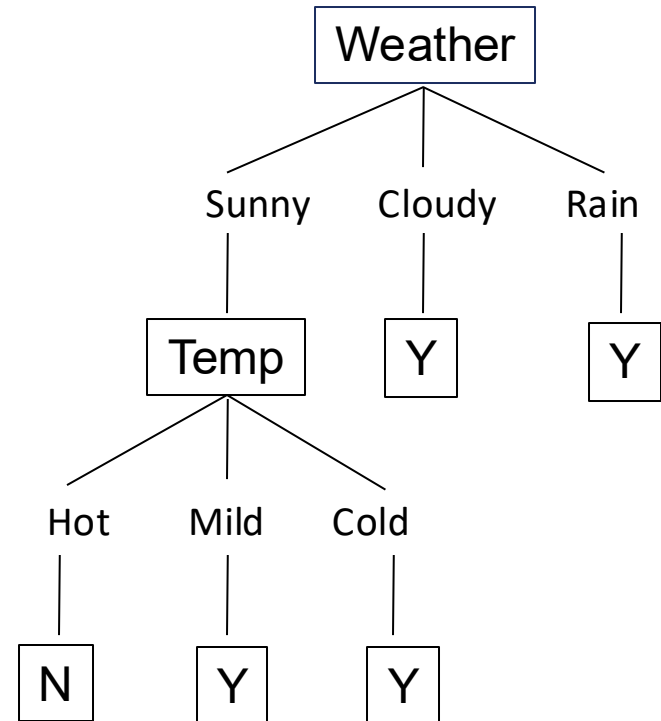Decision Tree Algorithm is used for both classification and regression task.

Example:

Decision Tree (ALGORITHM)

Decision Tree Algorithm is used for both classification and regression task.
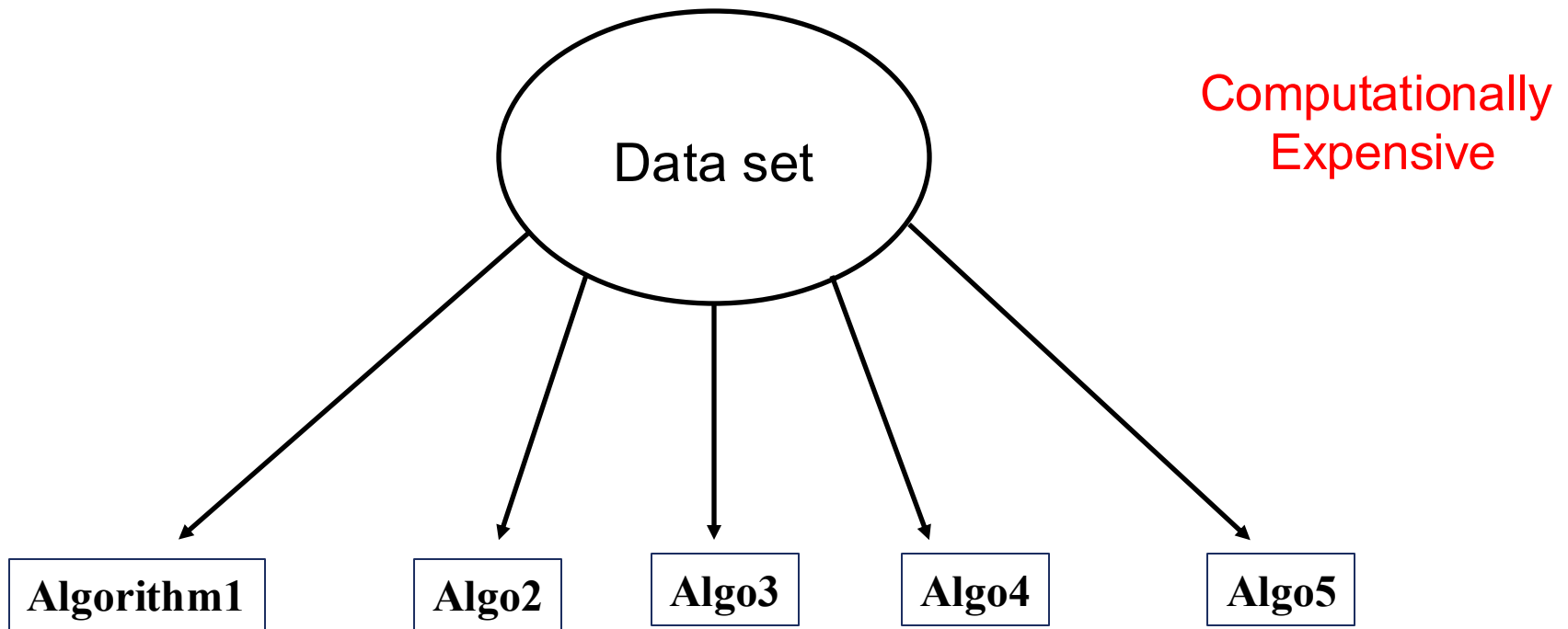
Example:

Few Key Attributes:

- Tree Structure
- Decision Nodes
- Leaf Nodes
- Splitting
- Entropy and Information Gain
- Pruning

```
                        Weather
              /            |          \
          Sunny         Cloudy        Rain
            |              |            |
          Temp            Y            Y
         /  |  \
      Hot  Mild  Cold
       |    |     |
       N    Y     Y
```

Ensemble Learning

Designed to make more accurate prediction

Data set

Computationally
Expensive

Algorithm1 **Algo2** **Algo3** **Algo4** **Algo5**

Final prediction is made based on majority voting

Types of Ensemble Learning

❑ Bootstrap Aggregating

❑ Boosting

❑ Stocking

❑ Voting

K-mean Clustering Method

Shopping pattern by customers

Customers	Age	Amount of shopping
C1	20	400
C2	40	700
C3	35	900
C4	25	300
C5	15	500
C6	30	1200
C7	50	1000

Hierarchical Clustering

Idea is to build a hierarchy amongst the data

Example: (a simple data)

1, 5, 8, 10, 19, 20

Random Forest Method

❑ Random forests or random decision forests is an ensemble learning method for classification, regression

❑ Random decision forests correct overfitting by a single decision tree to their training set

✓ Step1: Create a Bootstrap Dataset from the original data by randomly choosing subsets of data (repetition is allowed)

✓ Step2: Create randomized decision tree from bootstrap dataset

✓ Step3: Finally output of the random forest is the class selected by most trees

kNN Learning for Classification and Regression problems

Classification Problem

Q1: Suppose you have the following dataset with two features (X,Y) and the corresponding labels

Data point	X	Y	Label
1	2	3	A
2	3	4	A
3	5	6	B
4	7	8	B
5	10	10	A

Consider a new data point with $x_i=6.0$, $y_i=7.0$, use the kNN algorithm with k=3 to predict the label for this new data point.

kNN Learning for Classification and Regression problems

Regression Problem

Q1: Consider a dataset with a single feature (X) and corresponding target values (Y) for each X.

Data point	X	Y
1	2	3
2	3	4
3	5	6
4	7	8
5	10	10

Now, consider a new data point with x_i=7.0. Using kNN algorithm with K=2, predict the target value (Y) for this data point.