LECTURE 16

Classification

Building models of classification in sklearn

Data Science, Fall 2023 @ Knowledge Stream

Sana Jabbar

Outline

Lecture 16

- Introduction to Classification
- Types of Classification
- Classification Algorithms
- Performance Metrics
- Applications of Classification
- Overfitting and Underfitting
- Conclusion

Supervised Learning

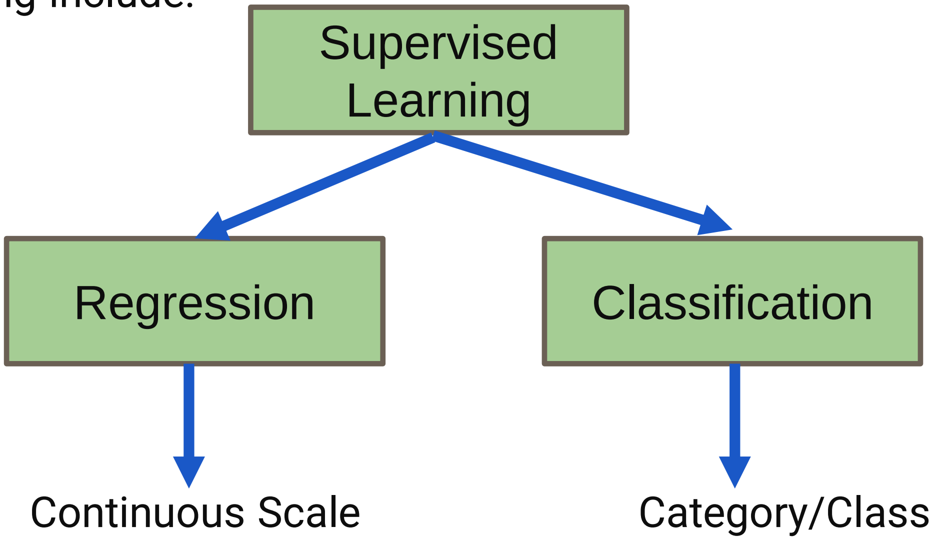
The model learns by example

Along with our input variable, we also give our model the corresponding correct labels

While training, the model can find patterns between our data and those labels

Some examples of Supervised Learning include:

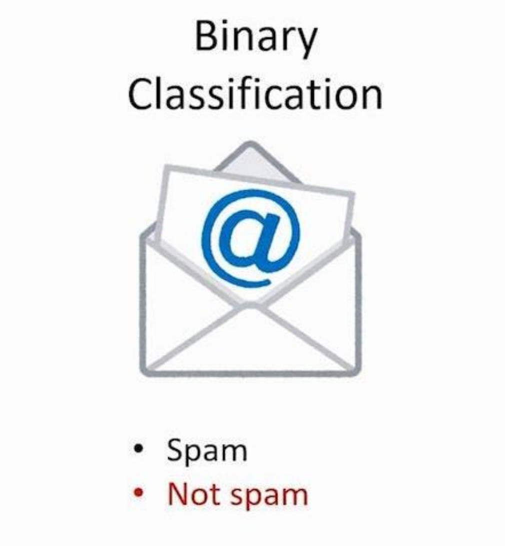
- Spam Detection
- Speech recognition
- Object Recognition



Classification

- Classification is defined as the process of recognition and grouping of objects
- Classification refers to a predictive modeling problem where a class label is predicted for a given example of input data
- For Classification, the training dataset must be sufficiently representative of the problem and have many examples of each class label.
- Types of classification

  1. Binary Classification



Classification

- Classification is defined as the process of recognition and grouping of objects
- Classification refers to a predictive modeling problem where a class label is predicted for a given example of input data
- For Classification, the training dataset must be sufficiently representative of the problem and have many examples of each class label.
- Types of classification

  1. Binary Classification

  2. Multi-Class Classification

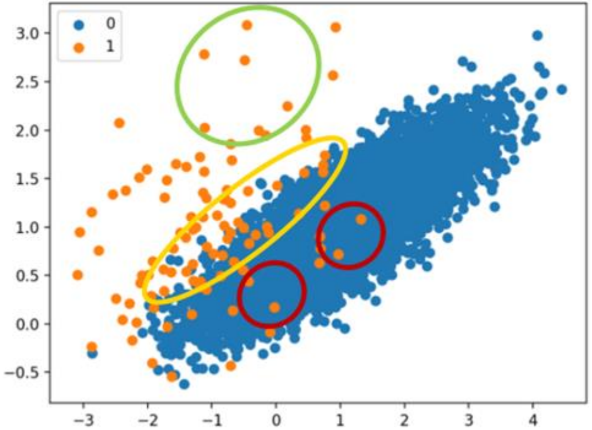
Classification

- Classification is defined as the process of recognition and grouping of objects
- Classification refers to a predictive modeling problem where a class label is predicted for a given example of input data
- For Classification, the training dataset must be sufficiently representative of the problem and have many examples of each class label.
- Types of classification

  1. Binary Classification
  2. Multi-Class Classification
  3. Multi-Label Classification

Classification

- Classification is defined as the process of recognition and grouping of objects
- Classification refers to a predictive modeling problem where a class label is predicted for a given example of input data
- For Classification, the training dataset must be sufficiently representative of the problem and have many examples of each class label.
- Types of classification
    1. Binary Classification
    2. Multi-Class Classification
    3. Multi-Label Classification
    4. Imbalanced Classification



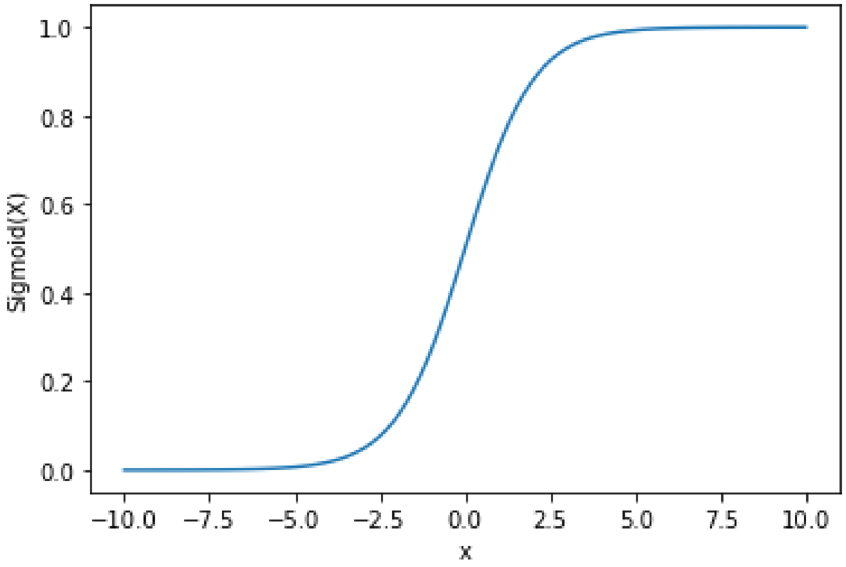
Binary Classification

- Refers to those classification tasks that have two class labels.
- Algorithms for binary classification
    1. Logistic Regression
    2. k-Nearest Neighbors
    3. Decision Trees
    4. Support Vector Machine
    5. Naive Bayes

Binary Classification

- Refers to those classification tasks that have two class labels.
- Algorithms for binary classification
    1. Logistic Regression
    2. k-Nearest Neighbors
    3. Decision Trees
    4. Support Vector Machine
    5. Naive Bayes

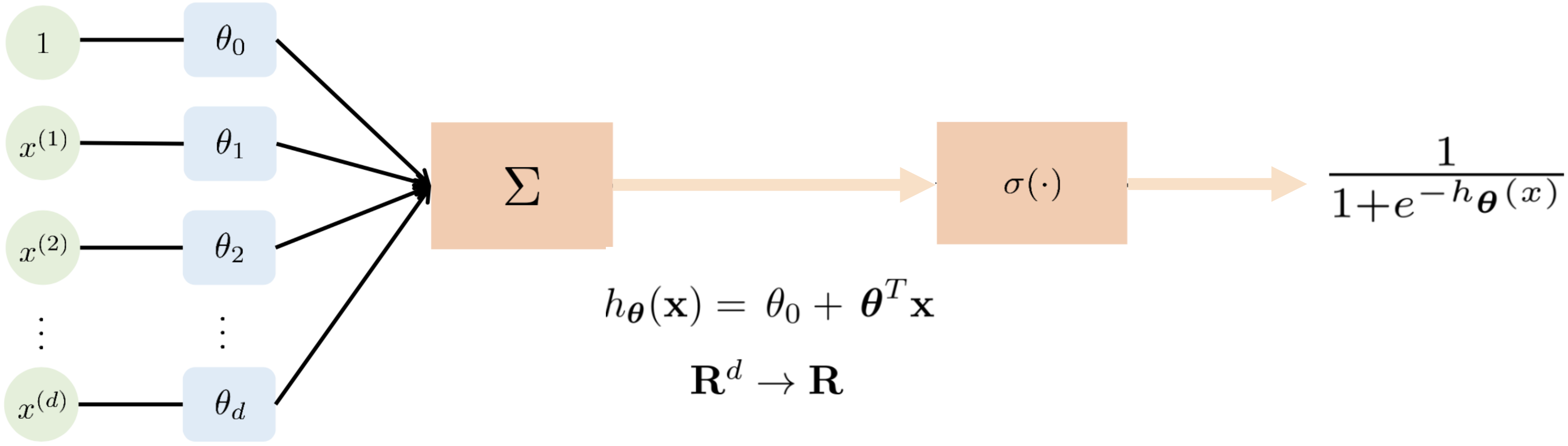
$$\mathrm{Sig}(x) = \frac{1}{1 + e^{-x}}$$



Linear Vs Logistic Regression

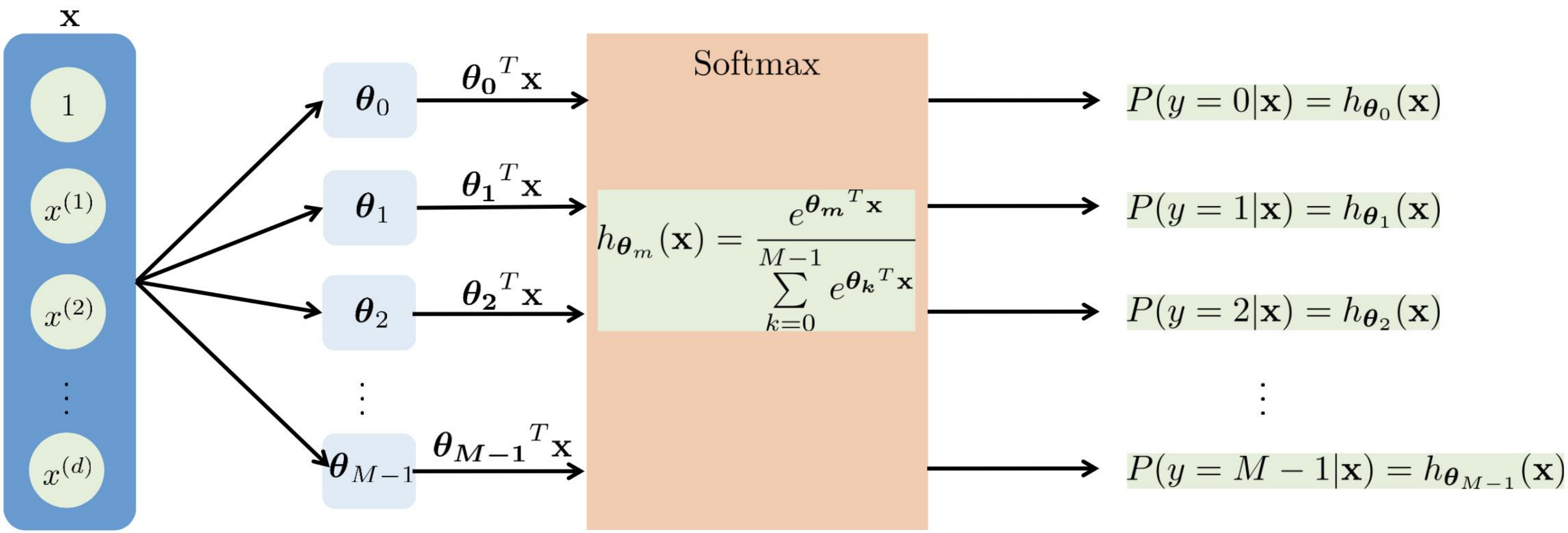
Linear Regression	Logistic Regression
The output is a continuous numeric value	The output is a probability value between 0 and 1
Uses linear combination of input features	Uses logistic function totransform the linear combination of input
$\mathrm{Y}_{\mathrm{pred}} = \theta^T X_{in}$	$\mathrm{Y}_{\mathrm{pred}} = \dfrac{1}{1 + e^{-\theta^T X_{in}}}$

Logistic Regression



$$h_{\boldsymbol{\theta}}(\mathbf{x}) = \theta_0 + \boldsymbol{\theta}^T \mathbf{x}$$

$$\mathbf{R}^d \to \mathbf{R}$$


$\mathbf{x}$
1
$x^{(1)}$
$x^{(2)}$
$x^{(d)}$
$\boldsymbol{\theta}_0$
$\boldsymbol{\theta}_1$
$\boldsymbol{\theta}_2$
$\boldsymbol{\theta}_{M-1}$
$\boldsymbol{\theta_0}^T\mathbf{x}$
$\boldsymbol{\theta_1}^T\mathbf{x}$
$\boldsymbol{\theta_2}^T\mathbf{x}$
$\boldsymbol{\theta_{M-1}}^T\mathbf{x}$
Softmax
$h_{\boldsymbol{\theta}_m}(\mathbf{x}) = \frac{e^{\boldsymbol{\theta_m}^T\mathbf{x}}}{\sum_{k=0}^{M-1} e^{\boldsymbol{\theta_k}^T\mathbf{x}}}$
$P(y=0|\mathbf{x}) = h_{\boldsymbol{\theta}_0}(\mathbf{x})$
$P(y=1|\mathbf{x}) = h_{\boldsymbol{\theta}_1}(\mathbf{x})$
$P(y=2|\mathbf{x}) = h_{\boldsymbol{\theta}_2}(\mathbf{x})$
$P(y=M-1|\mathbf{x}) = h_{\boldsymbol{\theta}_{M-1}}(\mathbf{x})$

Logistic Regression

- `from sklearn.linear_model import LogisticRegression`
- `from sklearn.cross_validation import train_test_split`

- `logreg = LogisticRegression()`
- `logreg.fit(X_train, y_train)`
- `y_pred = logreg.predict(X_test)`

Confusion matrix

A confusion matrix is a table that is used to define the performance of a classification algorithm

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

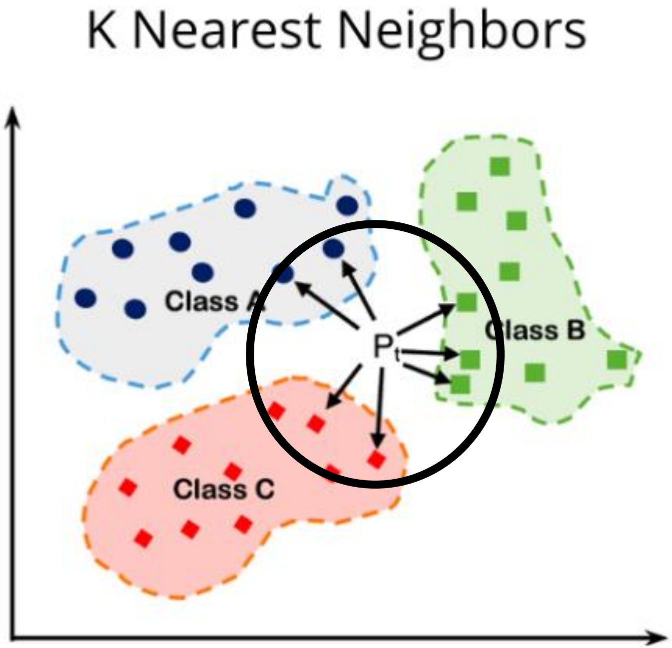
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{TP}{TP + \frac{1}{2}(FP + FN)}$$

	Predicted 0	Predicted 1
Actual 0	TN	FP
Actual 1	FN	TP

Binary Classification

- Refers to those classification tasks that have two class labels.
- Algorithms for binary classification
  1. Logistic Regression
  2. k-Nearest Neighbors
  3. Decision Trees
  4. Support Vector Machine
  5. Naive Bayes



KNN Algorithm

- The k-nearest neighbor algorithm stores all the available data
- Classifies a new data point based on the similarity measure (e.g., distance functions).
- The data point is classified by a majority vote of its neighbors, with the data point being assigned to the class most common amongst its K nearest neighbors measured by a distance function.

KNN Algorithm

- Loading the training and the test data.
- Choose the nearest data points (the value of K). K can be any integer.
- Do the following, for each test data point
  - Use Euclidean distance $\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2}$ or Manhattan distance $\sum_{i=1}^{k}|x_i - y_i|$
    to calculate the distance between test data and each row of training.
  - Sort the data set in ascending order based on the distance value.
  - From the sorted array, choose the top K rows.
  - Based on the most appearing class of these rows, it will assign a class to the test point.
  - End

Companies Using KNN

- Companies like Amazon or Netflix use KNN when recommending books to buy or movies to watch.
- How do these companies make recommendations?

Well, these companies gather data on the books you have read or movies you have watched on their website and apply KNN.

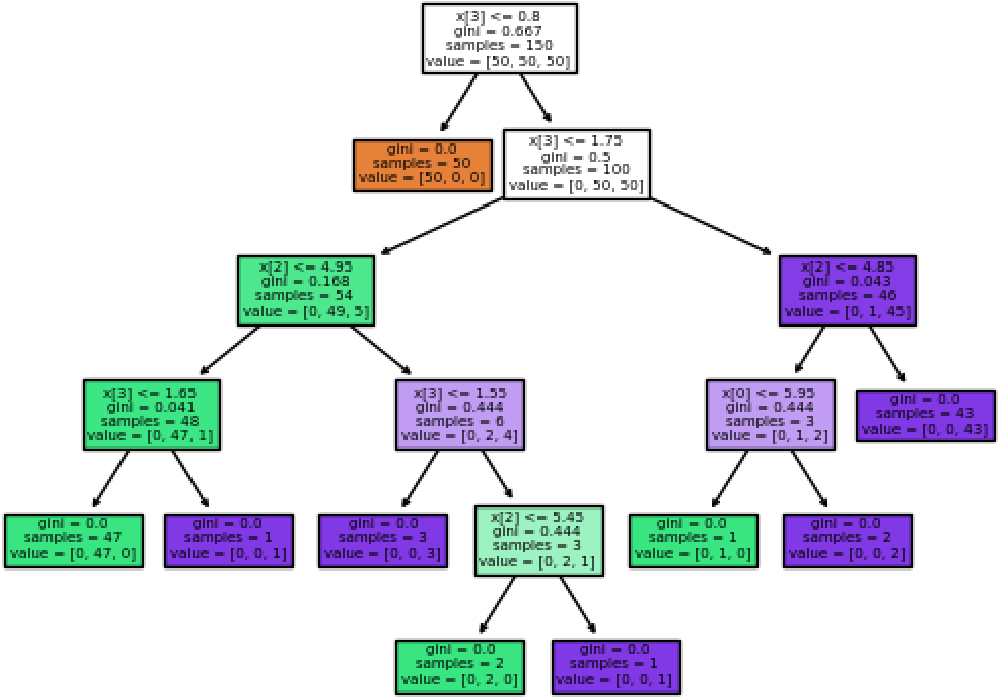
The companies will input your available customer data and compare that to other customers who have purchased similar books or have watched similar movies.

https://www.mygreatlearning.com/blog/knn-algorithm-introduction/

Decision Trees (DTs)

- Refers to those classification tasks that have two class labels.
- Algorithms for binary classification
    1. Logistic Regression
    2. k-Nearest Neighbors
    3. Decision Trees
    4. Support Vector Machine
    5. Naive Bayes

Decision tree trained on all the iris features



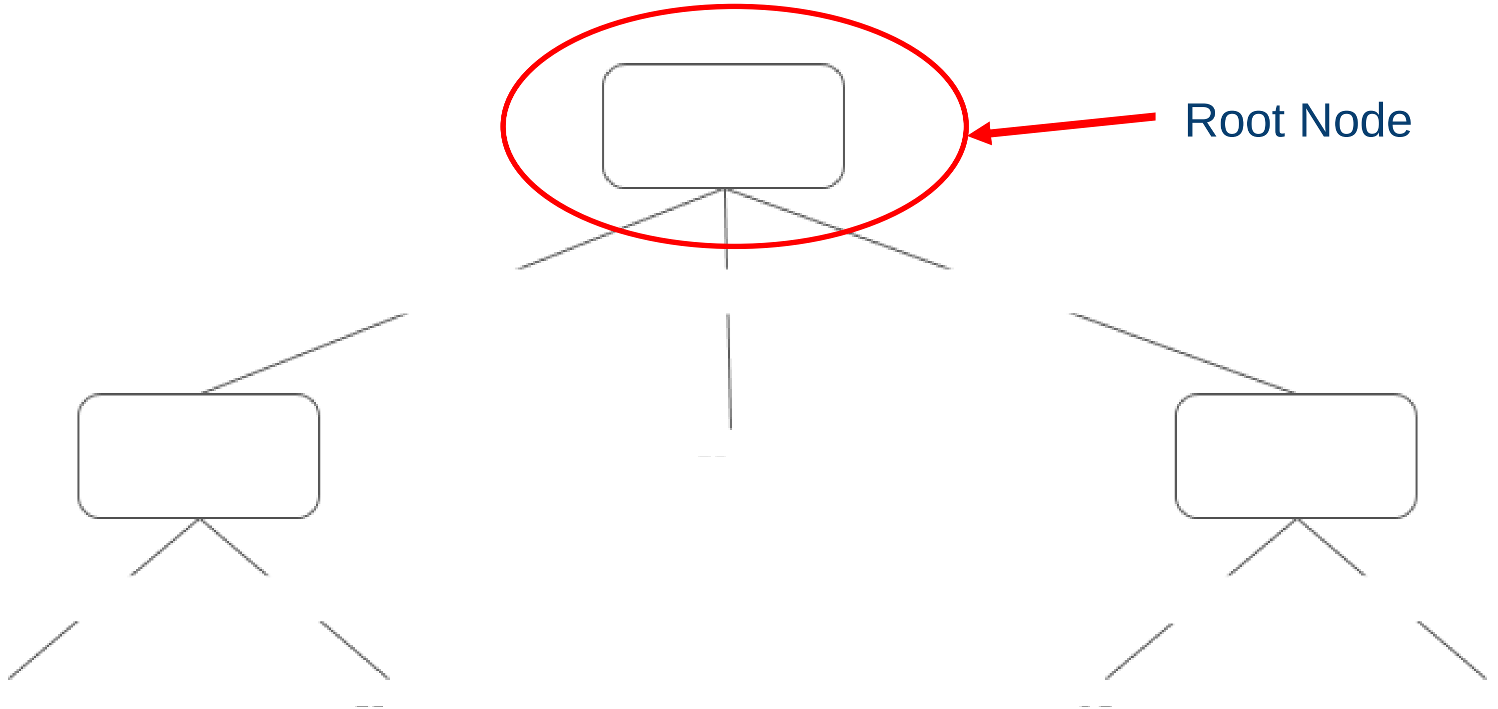
Decision Trees (DTs)

- A non-parametric supervised learning method used for Classification and regression.
- The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features

Decision Trees (DTs)

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Cloudy	Hot	High	Weak	Yes
4	Rainy	Mild	High	Weak	Yes
5	Rainy	Cool	Normal	Weak	Yes
6	Rainy	Cool	Normal	Strong	No
7	Cloudy	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Coll	Normal	Weak	Yes
10	Rainy	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Cloudy	Mild	High	Strong	Yes
13	Cloudy	Hot	Normal	Weak	Yes
14	Rainy	Mild	High	Strong	No

Decision Trees (DTs)



- Entropy, Information gain
- Step1: Calculate the Entropy of the whole dataset

$$S\{+9,-5\} = -\frac{9}{14}\log\left(\frac{9}{14}\right) - \frac{5}{14}\log\left(\frac{5}{14}\right) = 0.94$$

- Entropy of all attributes:

$$\text{Weather Sunny: } S\{+2,-3\} = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.97$$

$$\text{Weather Cloudy: } S\{+4,0\} = -\frac{4}{4}\log\left(\frac{4}{4}\right) - \frac{0}{4}\log\left(\frac{0}{4}\right) = 0$$

$$\text{Weather Rainy: } S\{+3,-2\} = -\frac{3}{5}\log\left(\frac{3}{3}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.97$$

Information Gain: Entropy (whole dataset) $-\frac{5}{14}\text{Ent(Sunny)} - \frac{4}{14}\text{Ent(Cloudy)} - \frac{5}{14}\text{Ent(Rainy)}$= 0.246

- Entropy, Information gain
- Step1: Calculate the Entropy of the whole dataset

$$S\{+9,-5\} = -\frac{9}{14}\log\left(\frac{9}{14}\right) - \frac{5}{14}\log\left(\frac{5}{14}\right) = 0.94$$

- Entropy of all attributes:

$$\text{Temp Hot: } S\{+2,-2\} = -\frac{2}{4}\log\left(\frac{2}{4}\right) - \frac{2}{4}\log\left(\frac{2}{4}\right) = \mathbf{1.0}$$

$$\text{Temp Mild: } S\{+4,-2\} = -\frac{4}{6}\log\left(\frac{4}{4}\right) - \frac{2}{6}\log\left(\frac{2}{6}\right) = 0.91$$

$$\text{Temp Cool: } S\{+3,-1\} = -\frac{3}{4}\log\left(\frac{3}{4}\right) - \frac{1}{4}\log\left(\frac{1}{4}\right) = 0.81$$

Information Gain: Entropy (whole dataset) $-\frac{4}{14}\mathrm{Ent(Hot)} - \frac{6}{14}\mathrm{Ent(Mild)} - \frac{4}{14}\mathrm{Ent(Coll)}$= 0.029

- Entropy, Information gain
- Step1: Calculate the Entropy of the whole dataset

$$S\{+9,-5\} = -\frac{9}{14}\log\left(\frac{9}{14}\right) - \frac{5}{14}\log\left(\frac{5}{14}\right) = 0.94$$

- Entropy of all attributes:

$$\text{Humidity High: } S\{+3,-4\} = -\frac{3}{7}\log\left(\frac{3}{7}\right) - \frac{4}{7}\log\left(\frac{4}{7}\right) = \mathbf{0.98}$$

$$\text{Humidity Normal: } S\{+6,-1\} = -\frac{6}{7}\log\left(\frac{6}{7}\right) - \frac{1}{7}\log\left(\frac{1}{7}\right) = 0.59$$

Information Gain: Entropy (whole dataset) $-\frac{7}{14}\mathrm{Ent(High)} - \frac{7}{14}\mathrm{Ent(Normal)} = 0.15$

- Entropy, Information gain
- Step1: Calculate the Entropy of the whole dataset

$$S\{+9,-5\} = -\frac{9}{14}\log\left(\frac{9}{14}\right) - \frac{5}{14}\log\left(\frac{5}{14}\right) = 0.94$$

- Entropy of all attributes:

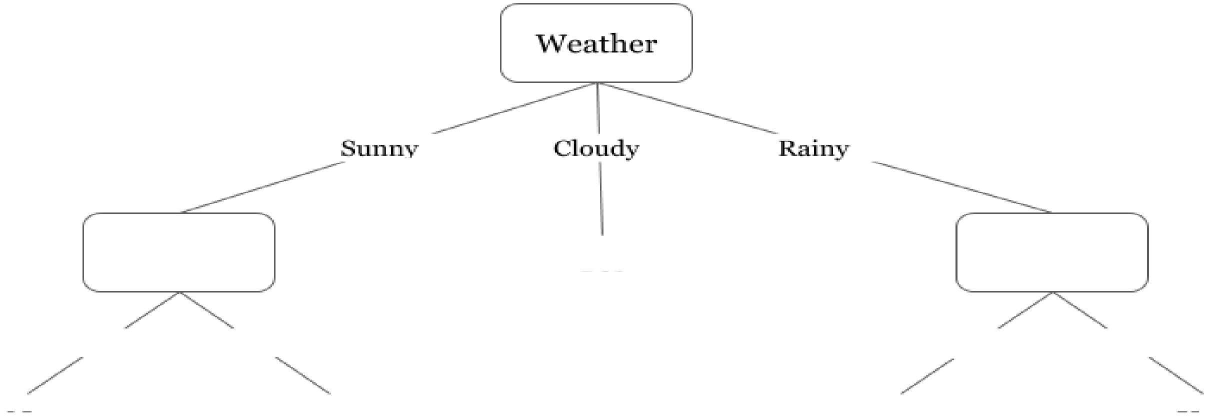
$$\text{Wind Weak: } S\{+3,-3\} = -\frac{3}{6}\log\left(\frac{3}{6}\right) - \frac{3}{6}\log\left(\frac{3}{6}\right) = 1.00$$

$$\text{Wind Strong : } S\{+6,-2\} = -\frac{6}{8}\log\left(\frac{6}{8}\right) - \frac{2}{8}\log\left(\frac{2}{8}\right) = 0.81$$

Information Gain: Entropy (whole dataset) $-\frac{6}{14}\mathrm{Ent(Weak)} - \frac{8}{14}\mathrm{Ent(Strong)} = 0.0478$

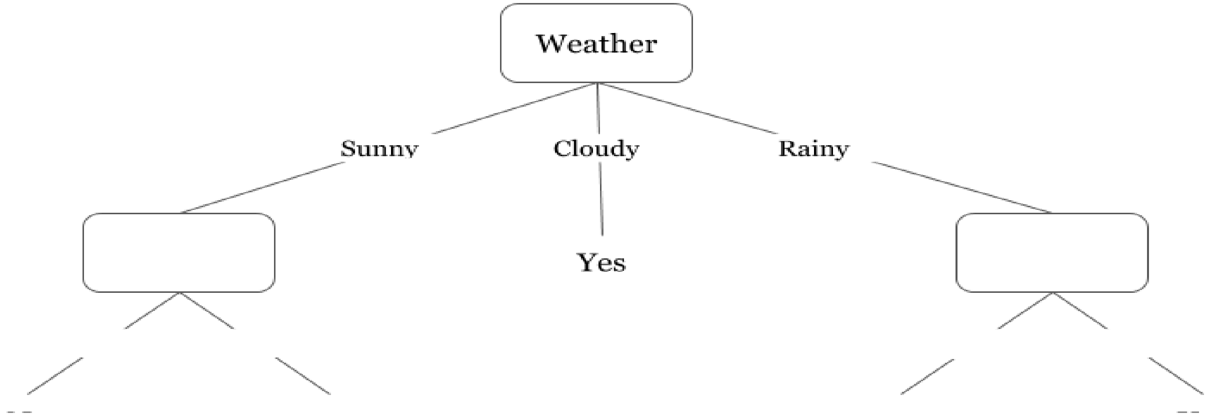
Decision Trees (DTs)

- Information Gain (S, Weather) = 0.246
- Information Gain (S, Temp) = 0.029
- Information Gain (S, Humidity) = 0.15
- Information Gain (S, Wind) = 0.0478



Decision Trees (DTs)

- Information Gain (S, Weather) = 0.246
- Information Gain (S, Temp) = 0.029
- Information Gain (S, Humidity) = 0.15
- Information Gain (S, Wind) = 0.0478



Decision Trees (DTs)

Day	Weather	Temperature	Humidity	Wind	Play?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes

- Entropy, Information gain
- Step1: Calculate the Entropy of Sunny

  S{+2,-3} = $-\frac{2}{5}\log\left(\frac{2}{5}\right)-\frac{3}{5}\log\left(\frac{3}{5}\right)=$ 0.97
- Entropy of all attributes:

  Temp Hot: S{0,-2} = $\mathbf{0}$

  Temp Mild: S{+1,-1} = $-\frac{1}{2}\log\left(\frac{1}{2}\right)-\frac{1}{2}\log\left(\frac{1}{2}\right)=$1.0

  Temp Cool: S{+,-0} = $\mathbf{0}$

Information Gain: Entropy (Sunny) $-\frac{2}{5}\,\mathbf{Ent(Mild)}=\mathbf{0.57}$

- Entropy, Information gain
- Step1: Calculate the Entropy of Sunny

$$S\{+2,-3\} = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.97$$

- Entropy of all attributes:

Humidity High: S{0,-3} = $\mathbf{0}$

Humidity Normal: S{+2,-0} = $\mathbf{0}$

Information Gain: Entropy (Sunny) $-\frac{3}{5}\mathbf{Ent(High)} - \frac{2}{5}\mathbf{Ent(Normal)} = \mathbf{0.97}$

- Entropy, Information gain
- Step1: Calculate the Entropy of Sunny

$$S\{+2,-3\} = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.97$$

- Entropy of all attributes:

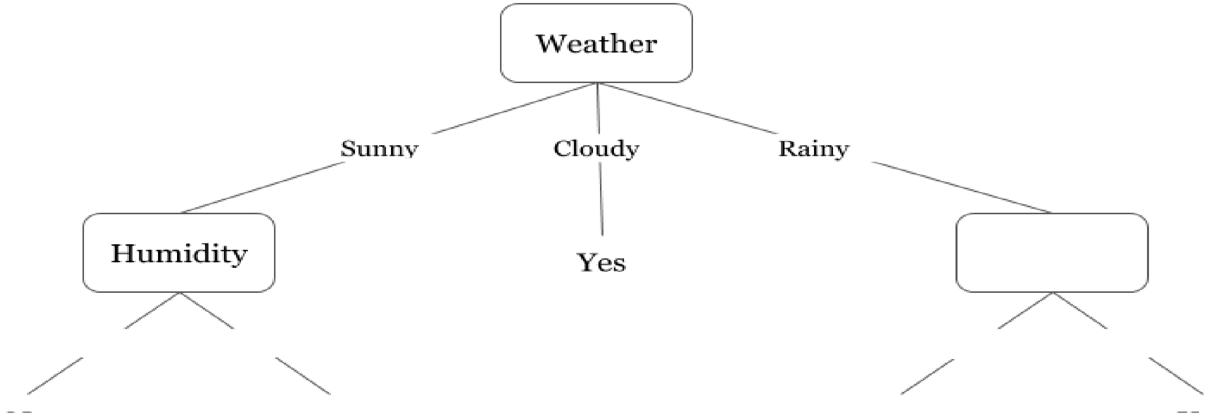
Wind Strong: $S\{+1,-1\} = \frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{1}{2}\log\left(\frac{1}{2}\right) = 1.0$

Wind Weak: $S\{+1,-2\} = \frac{1}{3}\log\left(\frac{1}{3}\right) - \frac{2}{3}\log\left(\frac{2}{3}\right) = 0.918$

Information Gain: Entropy (Sunny) $-\frac{2}{5}\mathrm{Ent}(\mathrm{Strong}) - \frac{2}{5}\mathrm{Ent}(\mathrm{Weak}) = 0.019$

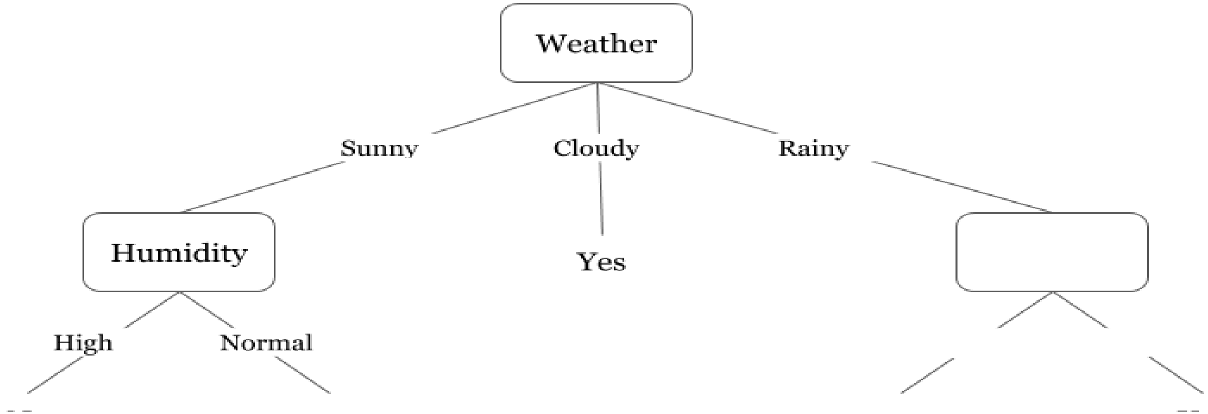
Decision Trees (DTs)

- Information Gain (S_sunny, Temp) = 0.57
- Information Gain (S_sunny, Humidity) = 0.97
- Information Gain (S_sunny, Wind) = 0.091



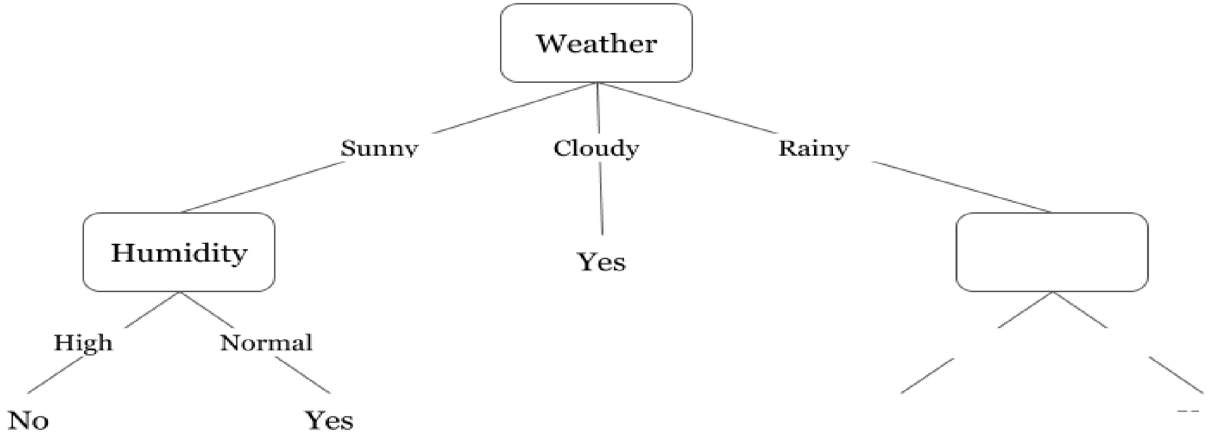
Decision Trees (DTs)

- Information Gain (S_sunny, Temp) = 0.57
- Information Gain (S_sunny, Humidity) = 0.97
- Information Gain (S_sunny, Wind) = 0.091



Decision Trees (DTs)

- Information Gain (S_sunny, Temp) = 0.57
- Information Gain (S_sunny, Humidity) = 0.97
- Information Gain (S_sunny, Wind) = 0.091



Decision Trees (DTs)

Day	Weather	Temperature	Humidity	Wind	Play?
4	Rainy	Mild	High	Weak	Yes
5	Rainy	Cool	Normal	Weak	Yes
6	Rainy	Cool	Normal	Strong	No
10	Rainy	Mild	Normal	Weak	Yes
14	Rainy	Mild	High	Strong	No

Decision Trees (DTs)

- Information Gain (S_rainy, Temp) = 0.019
- Information Gain (S_ rainy, Humidity) = 0.019
- Information Gain (S_ rainy, Wind) = 0.97