ML

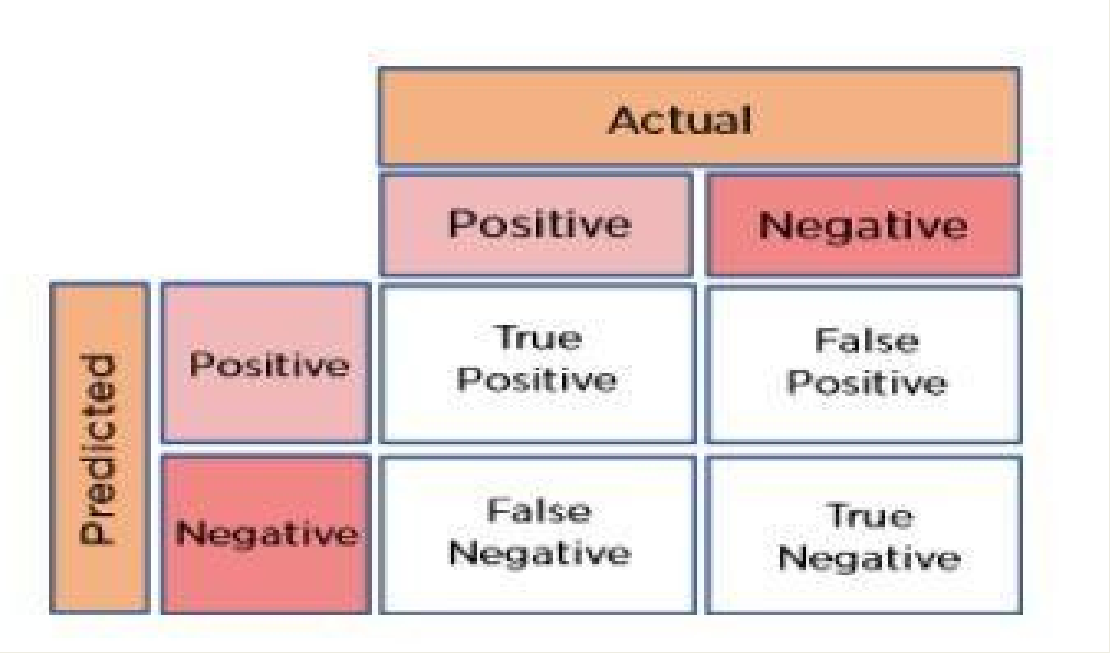
BY
Aradhana Behura
School Of Computer Engineering

F - measure (F1 Score) and Confusion Matrix

- **F1-Score: -** It is the **harmonic mean of precision and recall values for a classification problem**. The formula for F1-Score is as follows:

$$F_1 = \left(\frac{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}{2}\right)^{-1} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

- **Confusion matrix: -** It is an **N x N matrix used for evaluating the performance of a classification model, where N is the total number of target classes.** The matrix compares the actual target values with those predicted by the machine learning model.

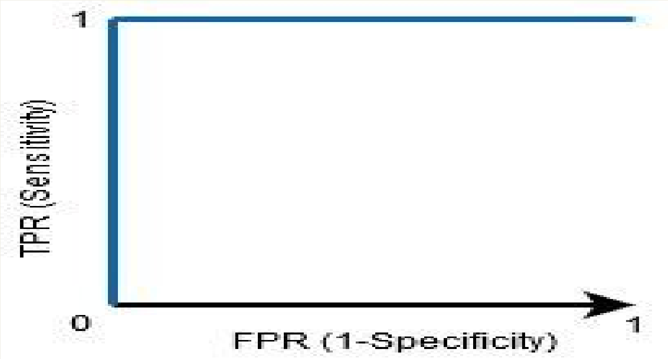


ROC curve

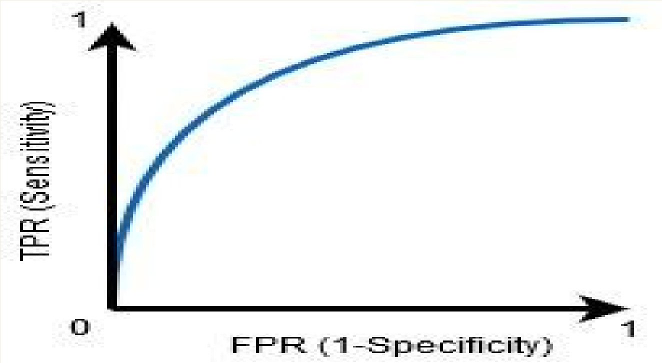
- **Receiver Operator Characteristic (ROC) curve: -** It is an **evaluation metric for binary classification problems.** It is a probability curve that **plots the TPR against FPR at various threshold values and essentially separates the 'signal' from the 'noise.'**

- In other words, it shows the performance of a classification model at all classification thresholds. The **Area Under the Curve (AUC) is the measure of the ability of a binary classifier to distinguish between classes and is used as a summary of the ROC curve.**

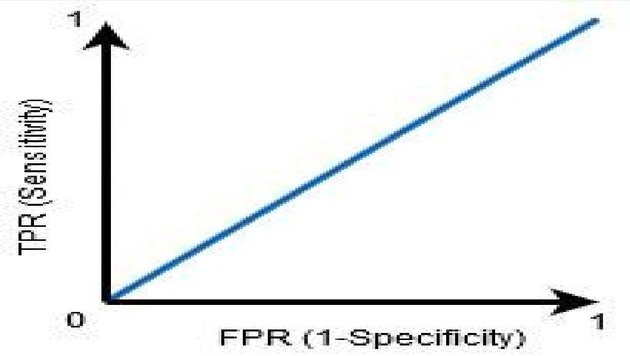
- **The higher the AUC, the better the model's performance** at distinguishing between the positive and negative classes.



When AUC = 1, the classifier can correctly distinguish between all the Positive and the Negative class points.



When 0.5<AUC<1, there is a high chance that the classifier will be able to distinguish the positive class values from the negative ones.



When AUC=0.5, then the classifier is not able to distinguish between Positive and Negative class points.

$$\text{TPR / Recall / Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

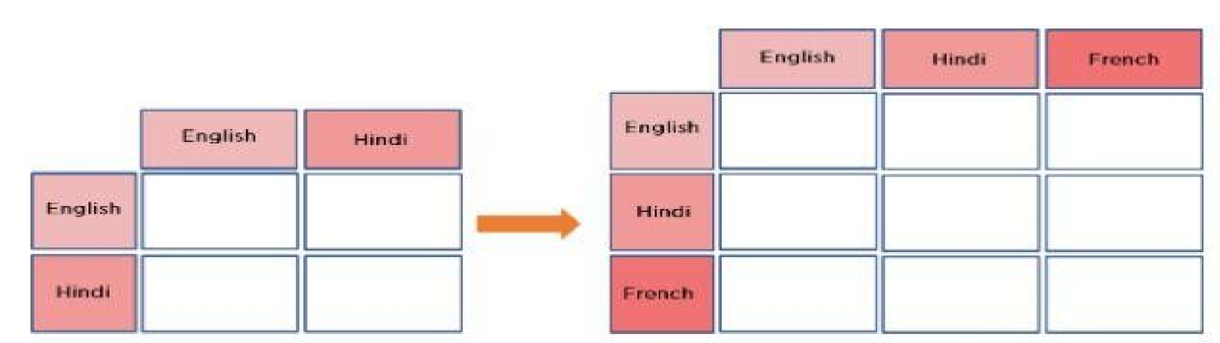
$$\text{FPR} = 1 - \text{Specificity} = \frac{FP}{TN + FP}$$

Solved problem on evaluation metrics

- Q. Consider a confusion matrix made for a classifier that classifies people based on whether they speak English or Spanish.

	English Speaker	Spanish Speaker
English Speaker	86	12
Spanish Speaker	10	79

- Ans:- From the above diagram, we can see that:
- True Positives (TP) = 86, True Negatives (TN) = 79, False Positives (FP) = 12, False Negatives (FN) = 10
- Accuracy = (86 +79) / (86 + 79 + 12 + 10) = 0.8823 = 88.23%
- Precision = 86 / (86 + 12) = 0.8775 = 87.75%
- Recall = 86 / (86 + 10) = 0.8983 = 89.83%
- F1-Score = (2* 0.8775 * 0.8983) / (0.8775 + 0.8983) = 0.8877 = 88.77%



To scale a confusion matrix, increase the number of rows and columns.
All the True Positives will be along the diagonal. The other values will be False Positives or False Negatives.

Different Metrics for comparing ML Algorithms

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

$$\text{Recall/ Sensitivity/True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

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

$$\text{Error Rate} = 1 - \text{Accuracy}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{F} - \text{measure/F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})}$$

$$\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

$$\text{True Negative Rate (TNR)} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{AUC ROC} = \frac{1}{2} \times \left(\frac{\text{TP}}{\text{TP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FP}}\right)$$

Naive Bayes Classifier

- Naïve Bayes algorithm is a supervised learning algorithm, which is **based on Bayes theorem and used for solving classification problems.**
- It is **mainly used in text classification that includes a high-dimensional training dataset.**
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a **probabilistic classifier, which means it predicts on the basis of the probability of an object.**
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles.**

- **Why is it called Naïve Bayes?**
- **Naïve:** It is called Naïve because it assumes that the **occurrence of a certain feature is independent of the occurrence of other features.** Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- **Bayes:** It is called Bayes because **it depends on the principle of Bayes' Theorem.**

Bayes' Theorem

- Bayes' theorem is also known as Bayes' Rule or Bayes' law, which is used to **determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.**
- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- Where,
- **P(A|B) is Posterior probability:** Probability of hypothesis A on the observed event B.
- **P(B|A) is Likelihood probability**: Probability of the evidence given that the probability of a hypothesis is true.
- **P(A) is Prior Probability:** Probability of hypothesis before observing the evidence.
- **P(B) is Marginal Probability:** Probability of Evidence.

Working of Naïve Bayes' Classifier

- Working of Naïve Bayes' Classifier can be understood with the help of the below example:
- Suppose we have a dataset of weather conditions and corresponding target variable "Play". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions. So to solve this problem, we need to follow the below steps:

- Convert the given dataset into frequency tables.
- Generate Likelihood table by finding the probabilities of given features.
- Now, use Bayes theorem to calculate the posterior probability.

- **Problem: If the weather is sunny, then the Player should play or not?**
- **Solution:** To solve this, first consider the dataset given right side:

	Outlook	Play
0	Rainy	Yes
1	Sunny	Yes
2	Overcast	Yes
3	Overcast	Yes
4	Sunny	No
5	Rainy	Yes
6	Sunny	Yes
7	Overcast	Yes
8	Rainy	No
9	Sunny	No
10	Sunny	Yes
11	Rainy	No
12	Overcast	Yes
13	Overcast	Yes

Working of Naïve Bayes' Classifier

❖ Frequency table for the Weather Conditions:

Weather	Yes	No
Overcast	5	0
Rainy	2	2
Sunny	3	2
Total	10	4

❖ Likelihood table weather condition:

Weather	No	Yes	
Overcast	0	5	5/14= 0.35
Rainy	2	2	4/14=0.29
Sunny	2	3	5/14=0.35
All	4/14=0.29	10/14=0.71	

❖ **Applying Bayes' theorem:**

❖P(Yes|Sunny)= P(Sunny|Yes)*P(Yes)/P(Sunny)
❖P(Sunny|Yes)= 3/10= 0.3
❖P(Sunny)= 0.35
❖P(Yes)=0.71
❖So P(Yes|Sunny) = 0.3*0.71/0.35= **0.60**

❖P(No|Sunny)= P(Sunny|No)*P(No)/P(Sunny)
❖P(Sunny|NO)= 2/4=0.5
❖P(No)= 0.29
❖P(Sunny)= 0.35
❖So P(No|Sunny)= 0.5*0.29/0.35 = **0.41**

❖ **Since P(Yes|Sunny)>P(No|Sunny), Hence on a Sunny day, Player can play the game.**

Advantages, Disadvantages and Applications of Naïve Bayes Classifier

- **Advantages:-**
- Naïve Bayes is **one of the fast and easy ML algorithms** to predict a class of datasets.
- It can be used for **Binary as well as Multi-class Classifications.**
- It performs well in Multi-class predictions as compared to the other Algorithms.
- It is the **most popular choice for text classification problems.**

- **Disadvantages:-**
- Naive Bayes assumes that all features are independent or unrelated, so **it cannot learn the relationship between features.**

- **Applications:-**
- It is used for **Credit Scoring.**
- It is used in **medical data classification.**
- It can be used in real-time predictions because Naïve Bayes Classifier is an eager learner.
- It is used in **Text classification such as Spam filtering and Sentiment analysis.**

Types of Naïve Bayes Model

- There are **three types of Naive Bayes Model**, which are given below:

- **Gaussian:** The Gaussian model assumes that **features follow a normal distribution.** This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.

- **Multinomial:** The Multinomial Naïve Bayes classifier is used when the **data is multinomial distributed. It is primarily used for document classification problems**, it means a particular document belongs to which category such as Sports, Politics, education, etc. **The classifier uses the frequency of words for the predictors.**

- **Bernoulli:** The Bernoulli classifier works **similar to the Multinomial classifier, but the predictor variables are the independent Booleans variables.** Such as if a particular word is present or not in a document. This model is **also famous for document classification tasks.**


Thank you