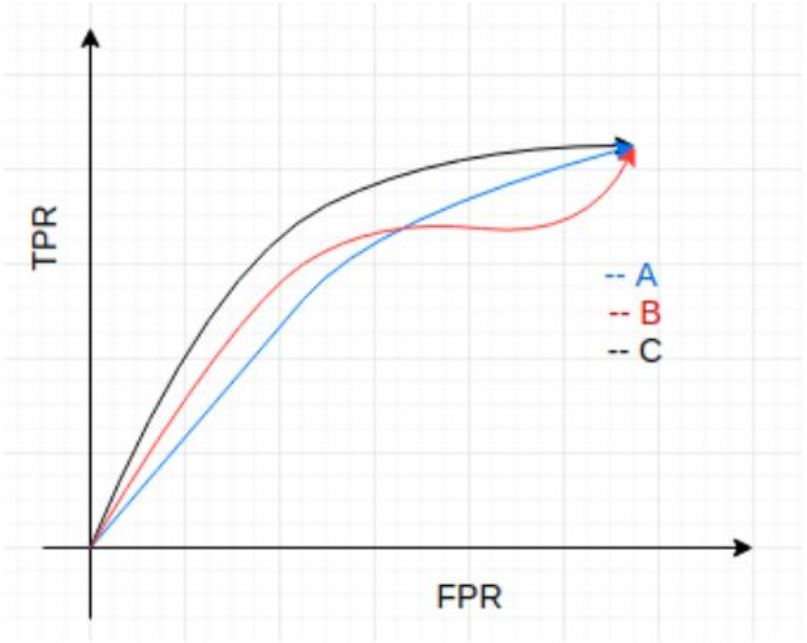
# TUTORIAL 10

Evaluating Classifiers and Clustering

1. **Consider the following confusion matrix and find the value of:**
    1. Precision
    2. Recall
    3. Accuracy
    4. Sensitivity
    5. Specificity

| | C1 | C2 |
|---|---|---|
| C1 | 10 | 3 |
| C2 | 5 | 16 |

## 2. The ROC curves for three classifiers A,B and C are shown:



**i) Out of A, B and C which is the best classifier?**

Ans. The classifier C dominates both A and B, therefore classifier C is best classifier.

**3. Run k-means clustering algorithm on following data points**

**(2,2) ,(4,2),(3,2),(2,3),(-1,1),(-2,0),(-1,-1) for k=2.**

*Initialize following data point to cluster 1 (4,2),(3,2),(2,3),(-1,1) and others to cluster 2 and start.*

**4. i) Suppose you have a single cluster of data points . The data points are (-2,-2),(-1,-2),(2,1),(1,2). Find the data point x which has highest average l2 distance with respect to other data points.**

**4. ii) Now we have two clusters , C1 with only x and C2 with other data points. For all the data points y in C2 , calculate it's distance from centroid of C1 and C2 and assign them to appropriate cluster. [Note the centroids are changing after moving each datapoint from C2 to C1] . What is final C1 and C2 you obtain ?**

## 4. iii) What type of clustering you just did ?

**5. Can you get different clusters depending on how you initialise the initial cluster ?**

**6) Which of the following statements is NOT TRUE about K- means clustering?**

a) It is an unsupervised learning algorithm

b) Overlapping of clusters is allowed in k-means clustering

c) It is a hard-clustering technique

d) k is a hyper parameter

**7) The K-means algorithm performs poorly:**

a) when data has noise and outliers

b) for categorical data where defining mean is difficult

c) Both a and b

d) None of the above