Análisis ROC

AADC Informática Clínica

ROC

1. Receiver Operator Characteristic Curve (ROC Curve)
2. For Binary Classification
3. Powerful performance measure

Probabilities as output

- Decision tree or knn can output probability that instance belong to class
- Define probability threshold from which you decide patient to be sick

Probabilities as output

- Example:
  - Threshold: 50%
  - Output:
    - New patient: 70%  30% 
    - Class new patient: 
  - Avoid sending sick patient home: lower threshold to 30%

Probabilities as output

- Threshold: 30%
  - More patients  classified as 
  - More patients  classified as 

ROC Curve

- Based on the confusion matrix
  - Binary classifier: positive (sick) or negative
  - Ratios:
    - True positive rate (TPR)= recall
    - False positive rate (FPR)

ROC Curve

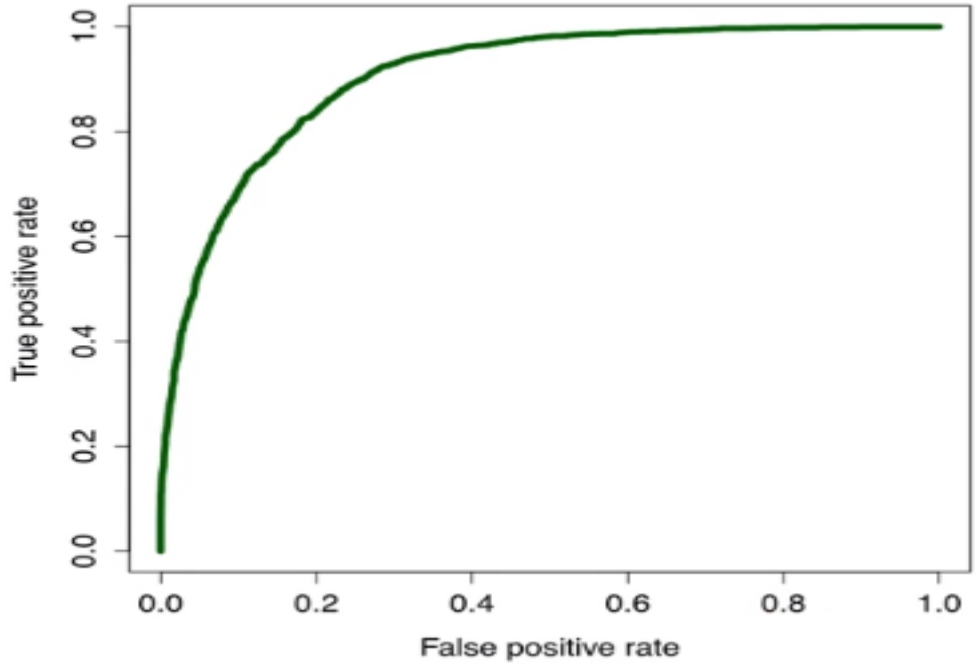
- Based on the confusion matrix
  - Binary classifier: positive (sick) or negative
  - Ratios:
    - True positive rate (TPR)= recall
      - TP / (TP + FN)
      - Truly sick patients that are diagnosed correctly as sick patients
    - False positive rate (FPR)

ROC Curve

- Based on the confusion matrix
  - Binary classifier: positive (sick) or negative
  - Ratios:
    - True positive rate (TPR)= recall
      - TP / (TP + FN)
      - Truly sick patients that are diagnosed correctly as sick patients
    - False positive rate (FPR)
      - FP / (FP + TN)
      - Falsy classified a negative (healthy) as positive (sick)

ROC Curve

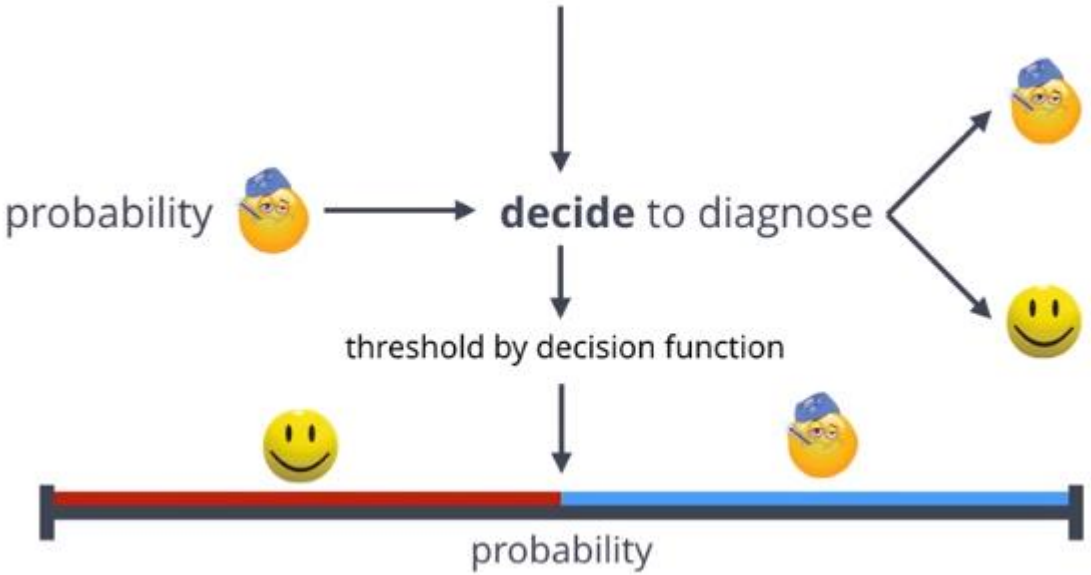
- Horizontal axis: FPR
- Vertical axis: TPR



ROC curve

- How to draw the curve?

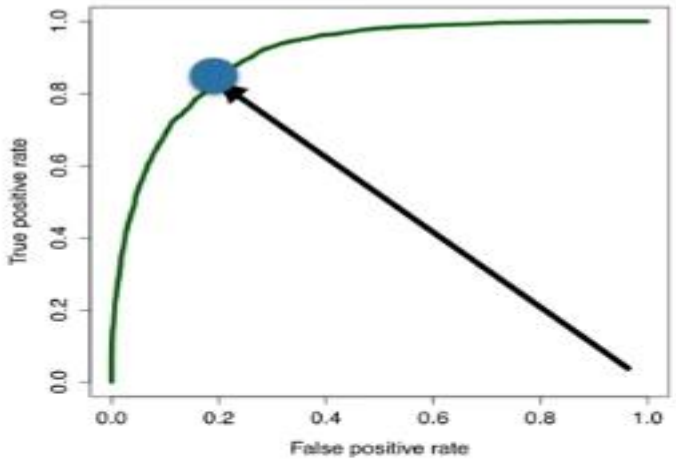
Decision function



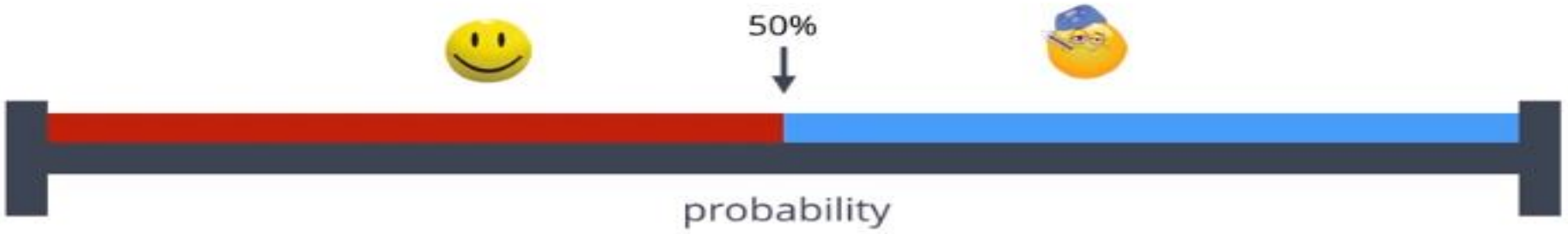
Source: Introduction to MachineLearning (Data camp)

ROC Curve

- How to draw the curve?

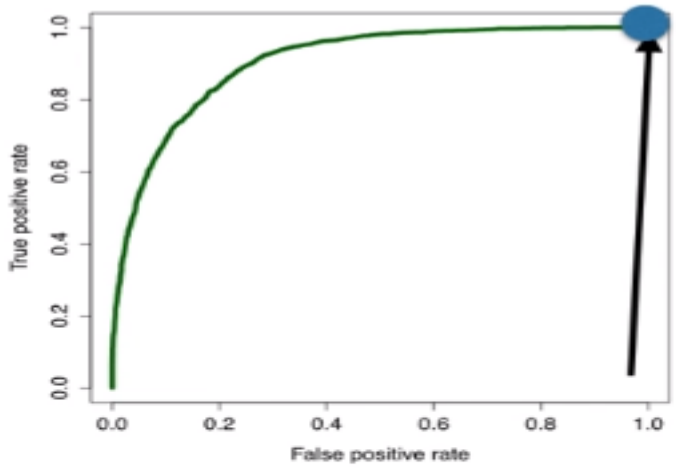


>=50%: sick
< 50%: healthy



Source: Introduction to MachineLearning (Data camp)

ROC Curve



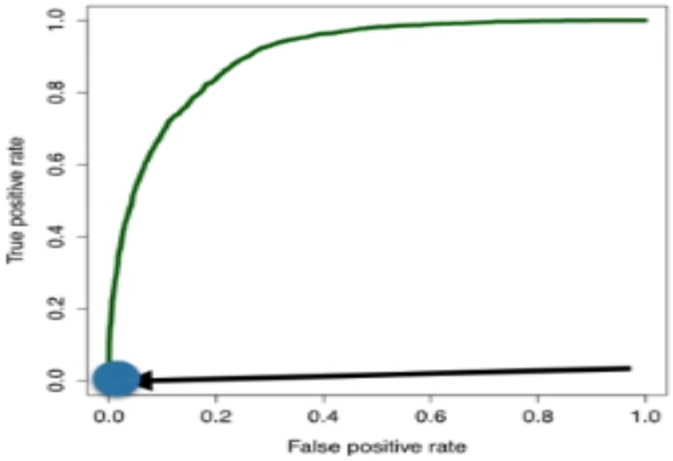
all sick

0%

probability

Source: Introduction to MachineLearning (Data camp)

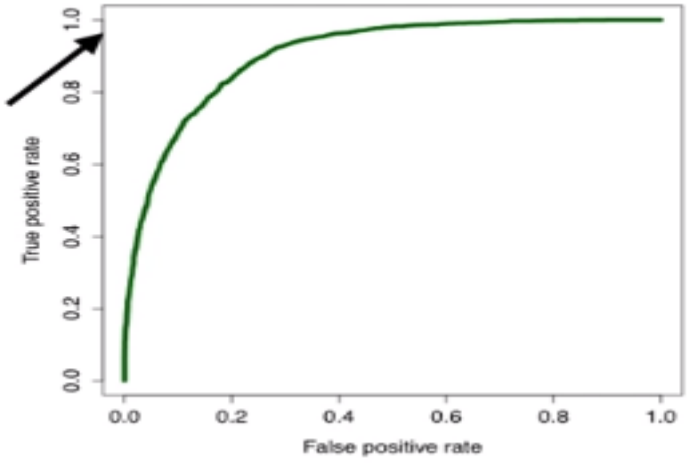
ROC Curve



Source: Introduction to MachineLearning (Data camp)

ROC Curve

- Is it a good curve?
- Good:



Introduction to MachineLearning (Data camp)

ROC Curve

- How to compare classifiers? == how to compare curves?
- Using AUC (area under the curve)