Train In Data

ROC Curve

Receiving Operating Characteristic (ROC)

- The ROC curve plots the benefits (TPrate) and costs (FPrate) at different classification thresholds.
- Every point on the ROC curve represents a probability threshold and the model performance trade-off
- Evaluates how well a classifier can separate positive and negative examples
- Helps identify the best threshold to separate them.

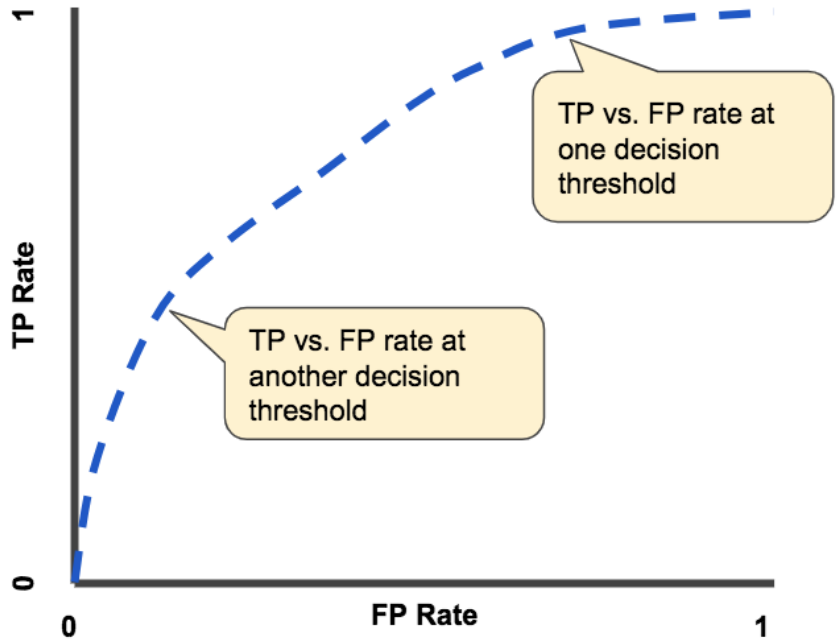


Figure 4. TP vs. FP rate at different classification thresholds.

Area Under the ROC: AUC

- AUC is the area under the ROC curve
- AUC provides an aggregate measure of performance across all possible classification thresholds.
- Higher AUC indicates the model is better at predicting both classes

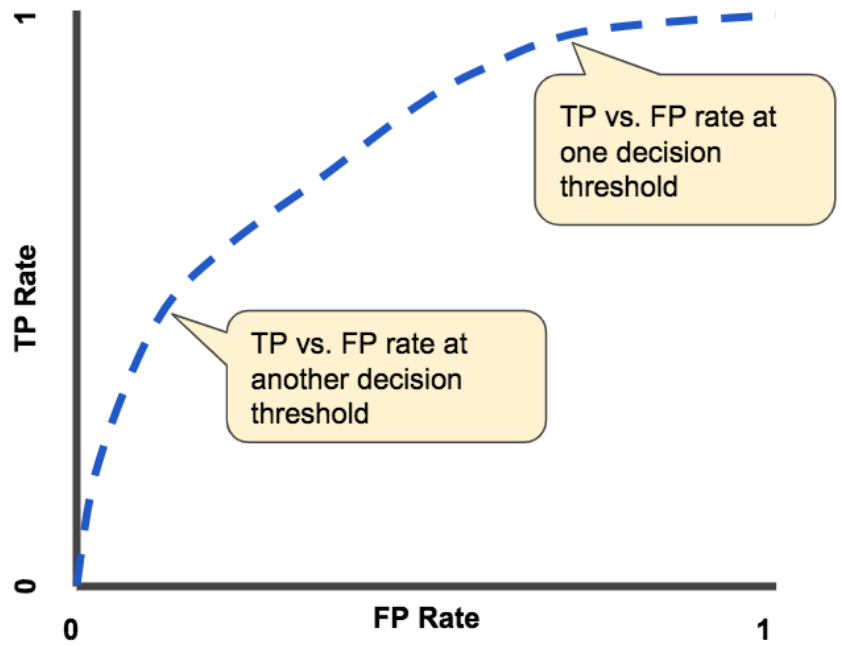
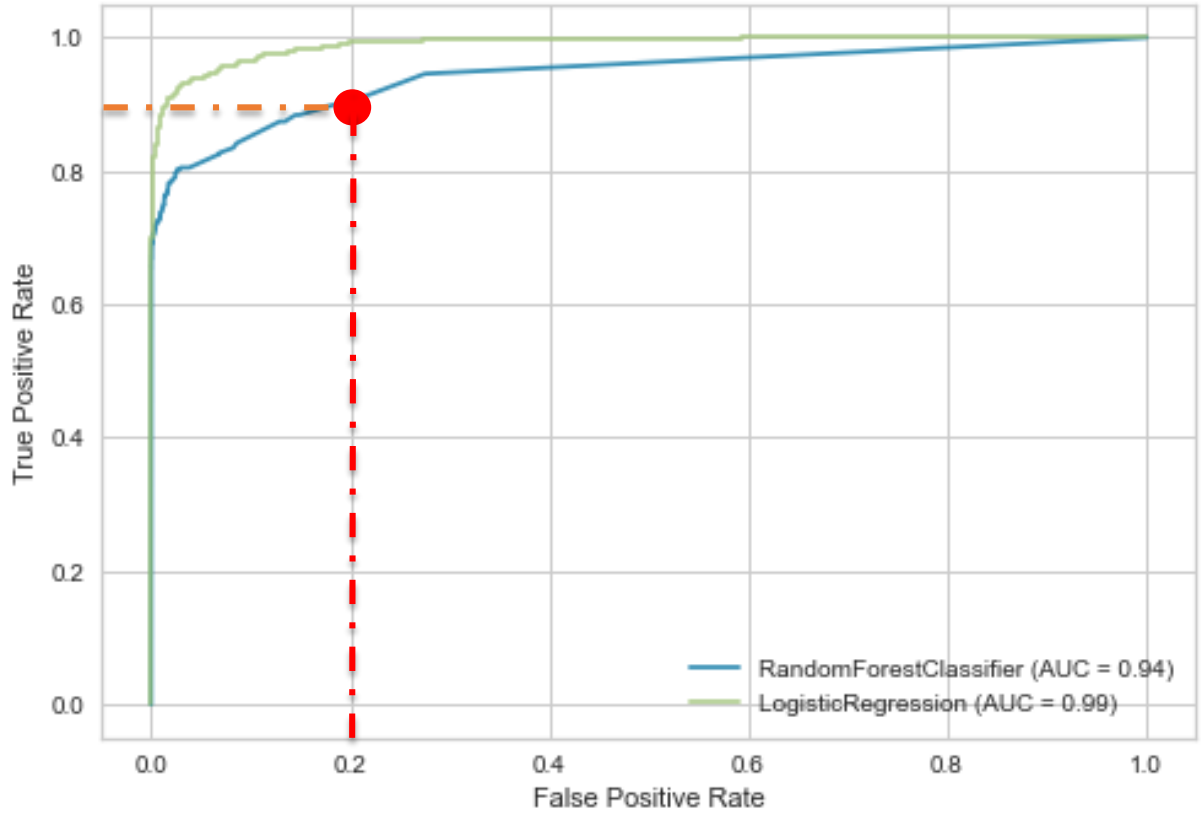


Figure 4. TP vs. FP rate at different classification thresholds.

ROC Curve

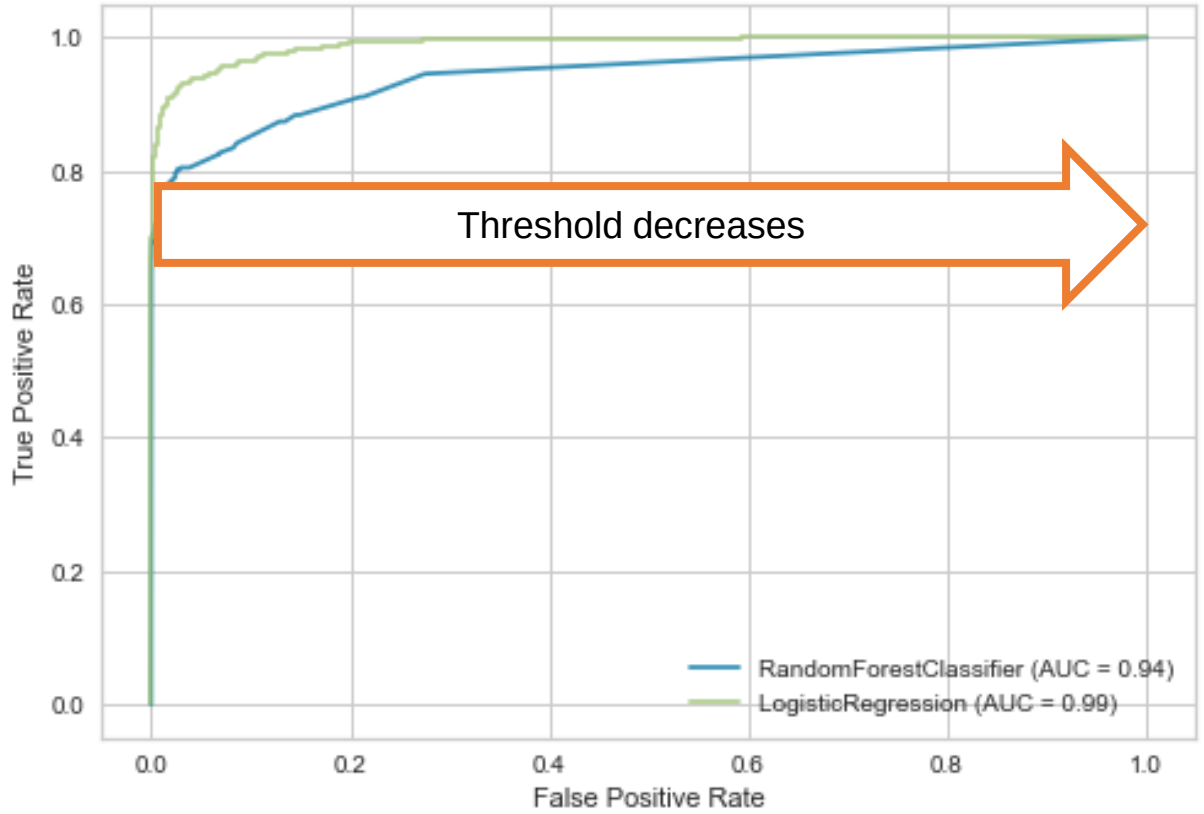


For the threshold at ● :

TPR = 0.9

FPR = 0.2

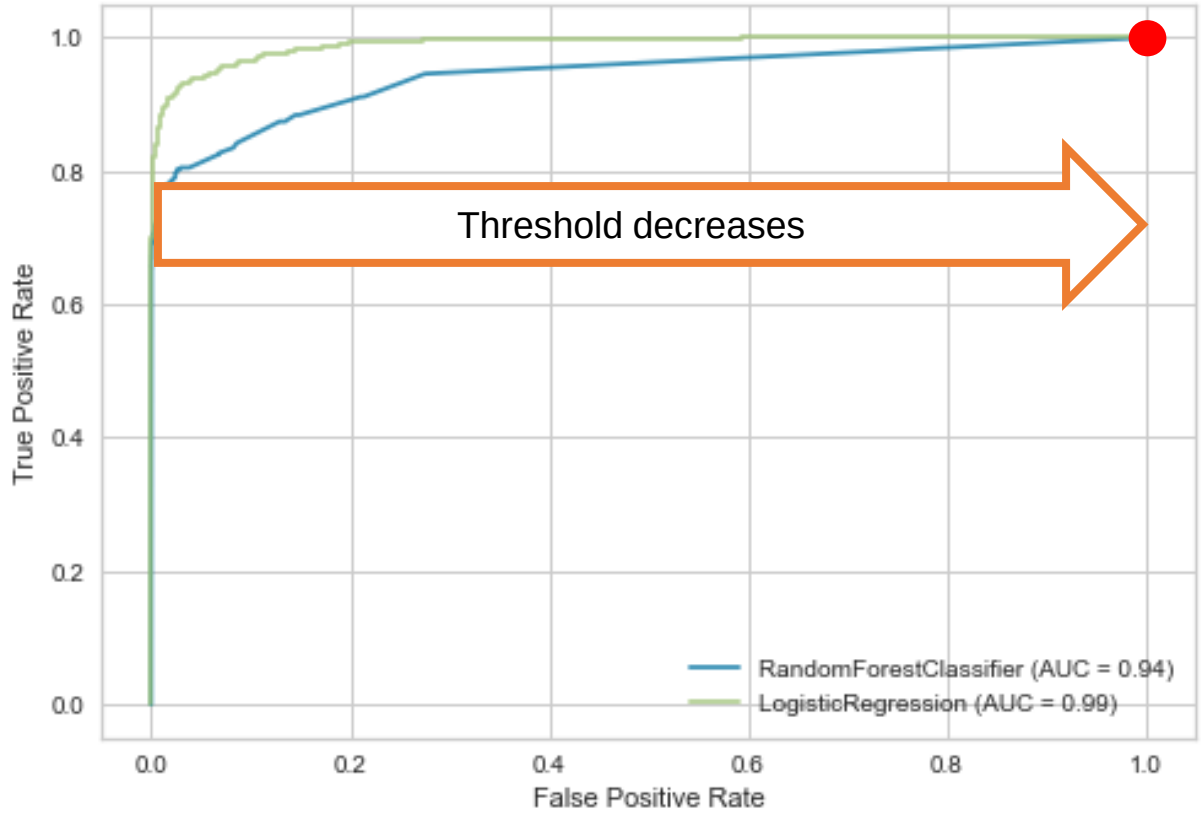
ROC Curve



As threshold decreases:

- TPR increases
  - TP / (TP + FP)
- FPR increases
  - FP / (FP + TN)

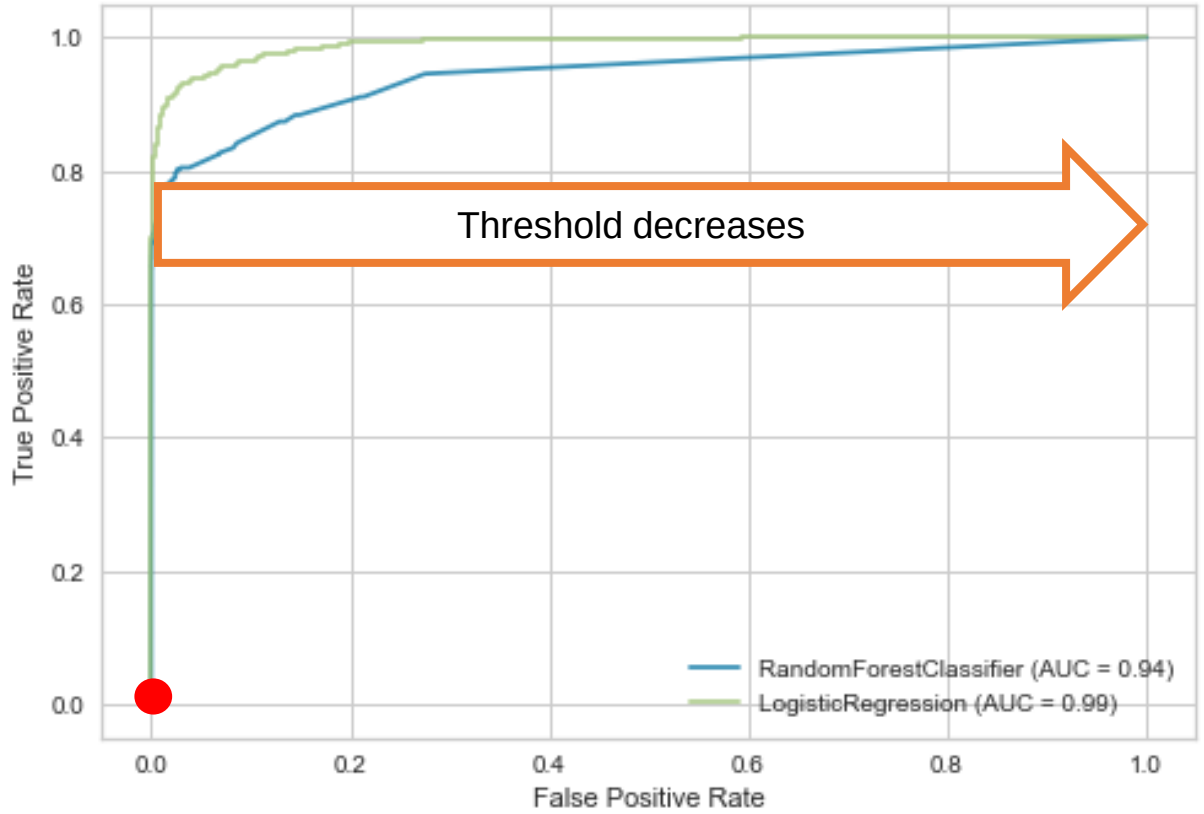
ROC Curve



As threshold decreases:

- TPR increases
  - TP / (TP + FP)
- FPR increases
  - FP / (FP + TN)
- Threshold = 0
  - Recall / TPR = 1
  - FPR ~ balancing ratio

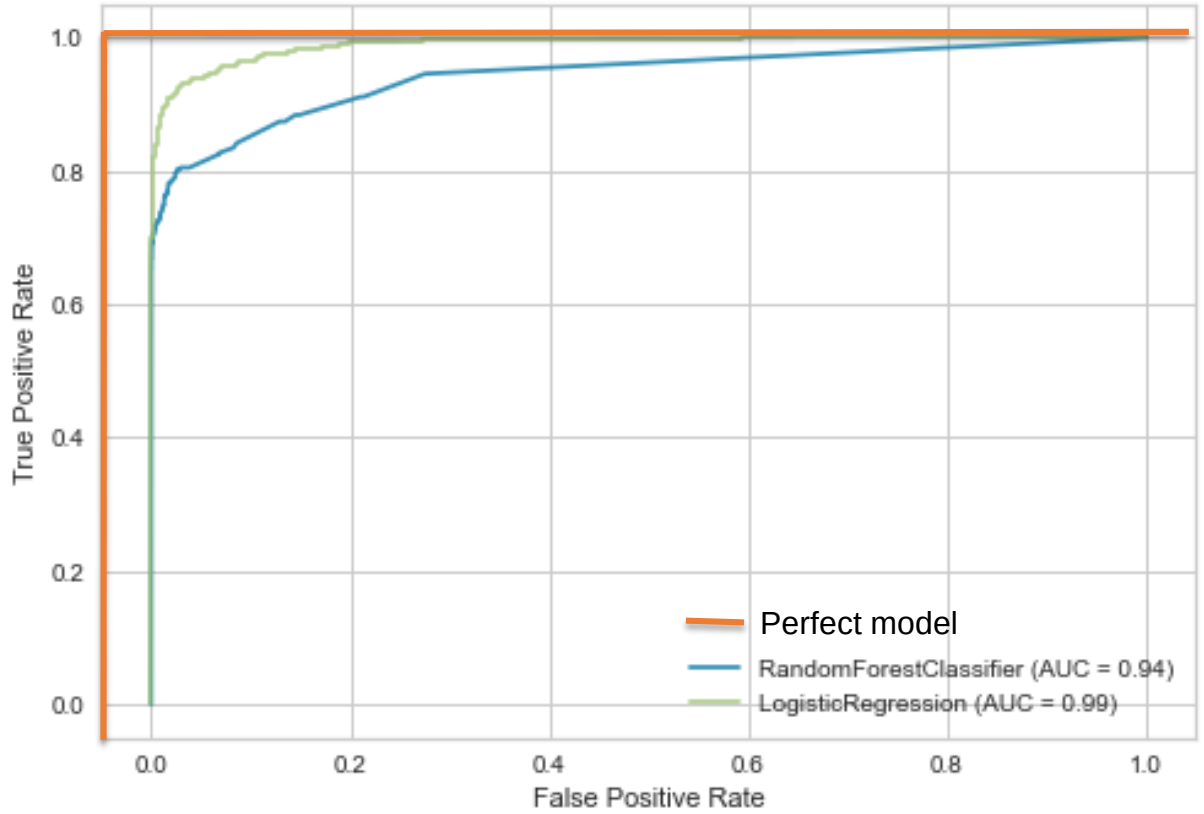
ROC Curve



As threshold decreases:

- TPR increases
  - TP / (TP + FP)
- FPR increases
  - FP / (FP + TN)
- Threshold approx. 1
  - Recall / TPR approx. 0
  - FPR approx. 0

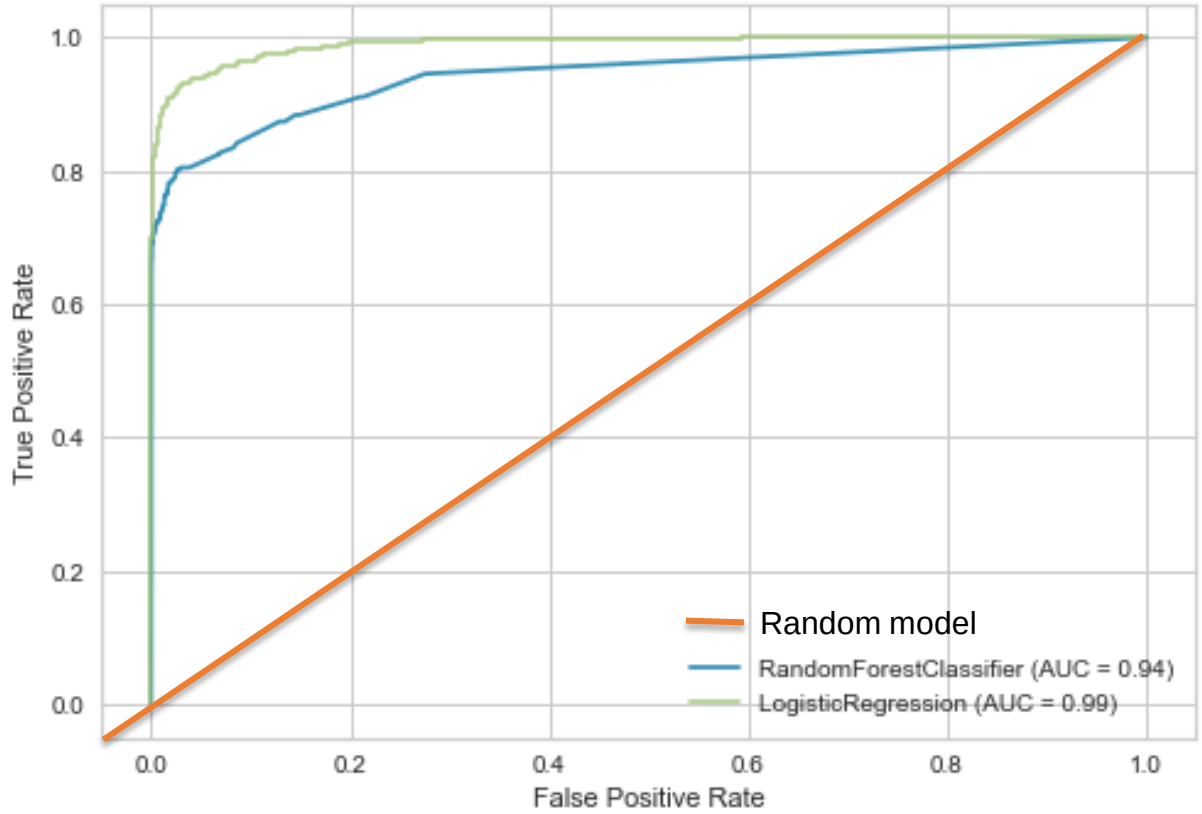
ROC: Perfect model



For the perfect model:

AUC = 1

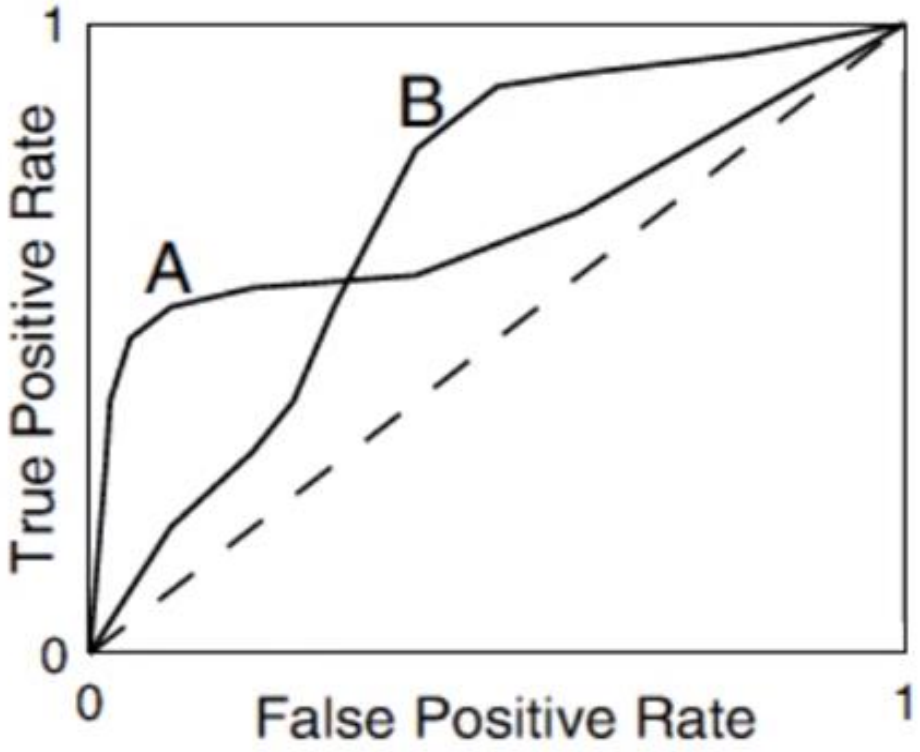
ROC: Random model



For a random model:

AUC = 0.5

Known issues of the ROC



ROC curves of different classifiers cross each other

The performance of the classifiers is not comparable for all discriminant thresholds

Train In Data

THANK YOU

www.trainindata.com