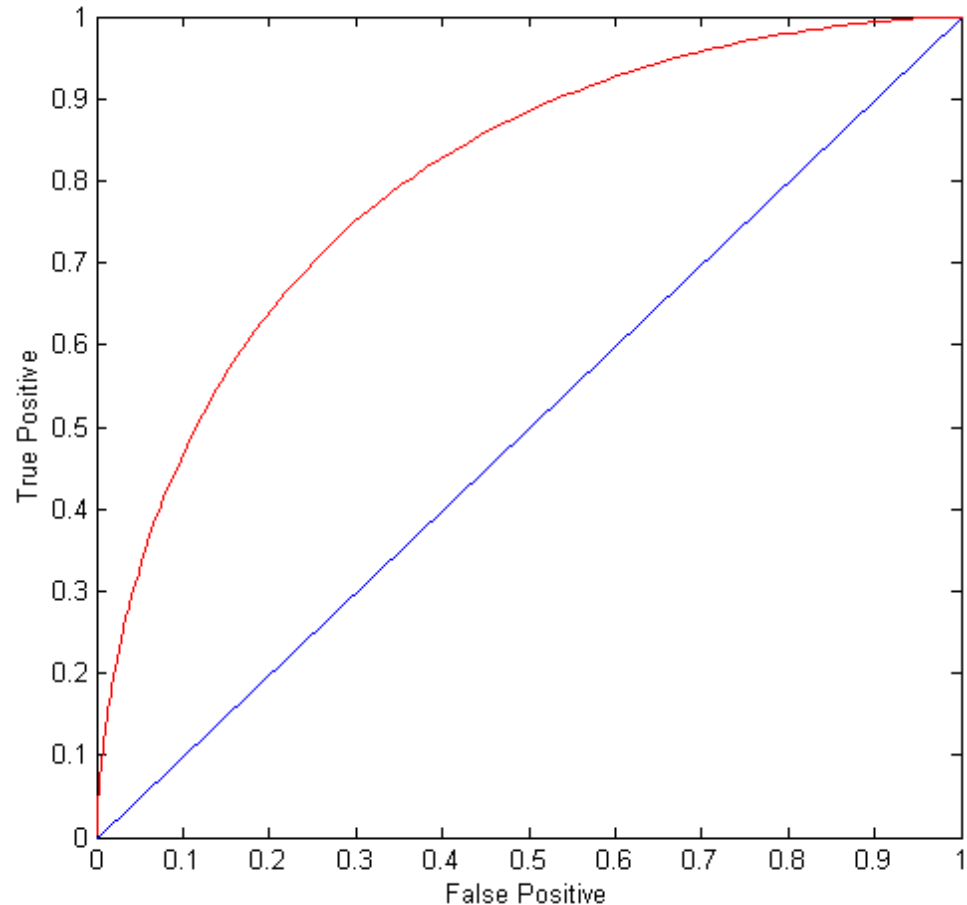# ROC (Receiver Operating Characteristic)

- A graphical approach for displaying trade-off between detection rate and false alarm rate

- Developed in 1950s for signal detection theory to analyze noisy signals

- ROC curve plots TPR against FPR
  - Performance of a model represented as a point in an ROC curve
  - Changing the threshold parameter of classifier changes the location of the point

ROC Curve

(TPR,FPR):

- ☐ (0,0): declare everything to be negative class

- ☐ (1,1): declare everything to be positive class

- ☐ (1,0): ideal

- ☐ Diagonal line:
  - − Random guessing
  - − Below diagonal line:
   - ◆ prediction is opposite of the true class

ROC (Receiver Operating Characteristic)

☐ To draw ROC curve, classifier must produce continuous-valued output

  – Outputs are used to rank test records, from the most likely positive class record to the least likely positive class record
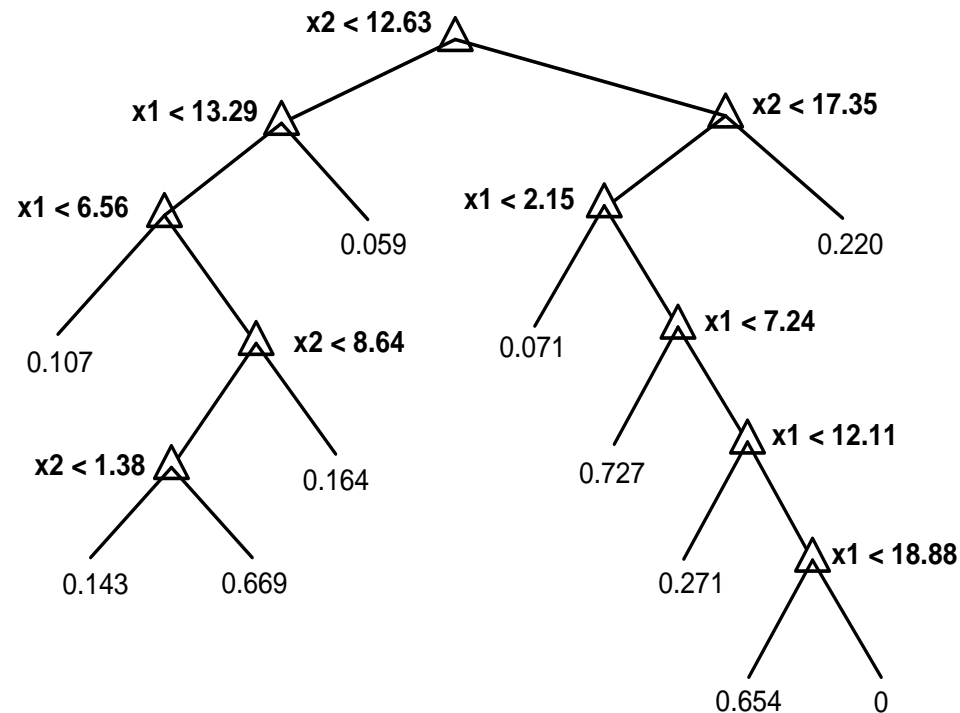
☐ Many classifiers produce only discrete outputs (i.e., predicted class)

  – How to get continuous-valued outputs?

    ◆ Decision trees, rule-based classifiers, neural networks, Bayesian classifiers, k-nearest neighbors, SVM
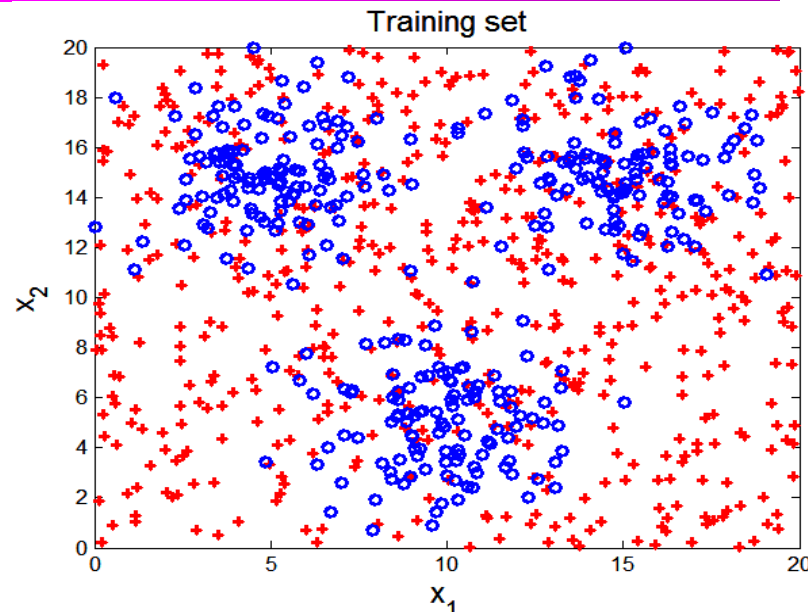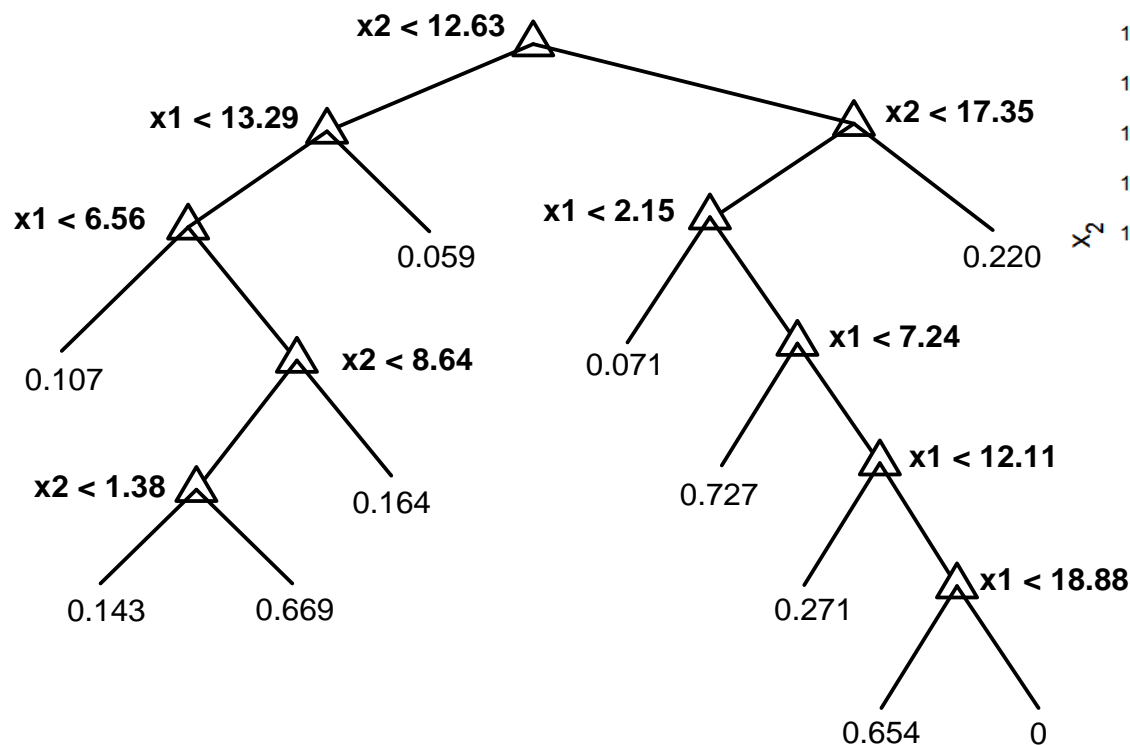
Example: Decision Trees

Decision Tree
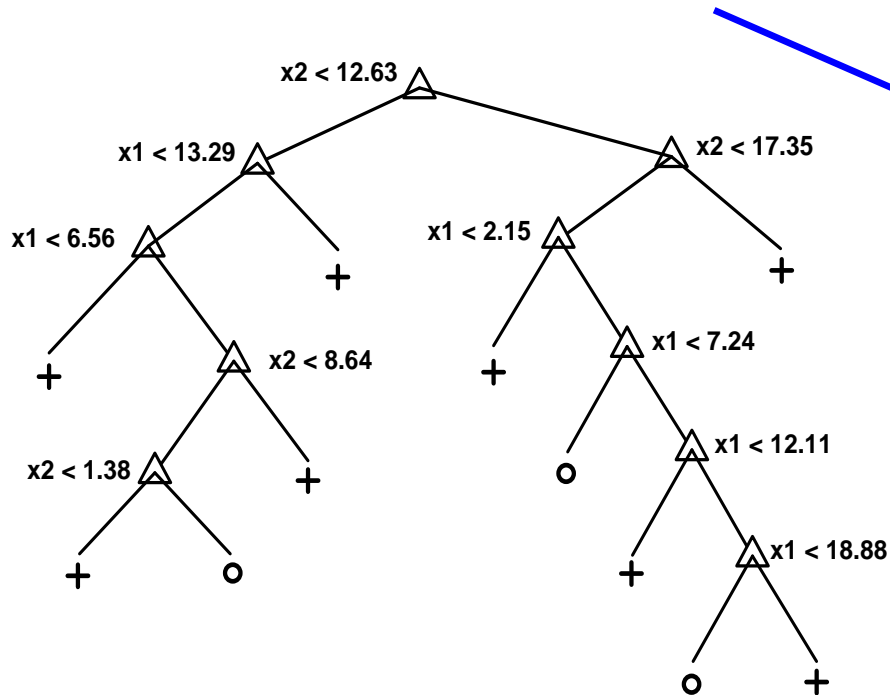
Continuous-valued outputs

ROC Curve Example

$\alpha = 0.3$		Predicted Class	
		Class o	Class +
Actual	Class o	645	209
Class	Class +	298	948

$\alpha = 0.7$		Predicted Class	
		Class o	Class +
Actual	Class o	181	673
Class	Class +	78	1168

ROC Curve Example

- 1-dimensional data set containing 2 classes (positive and negative)

- Any points located at x > t is classified as positive

At threshold t:

TPR=0.5, FNR=0.5, FPR=0.12, TNR=0.88
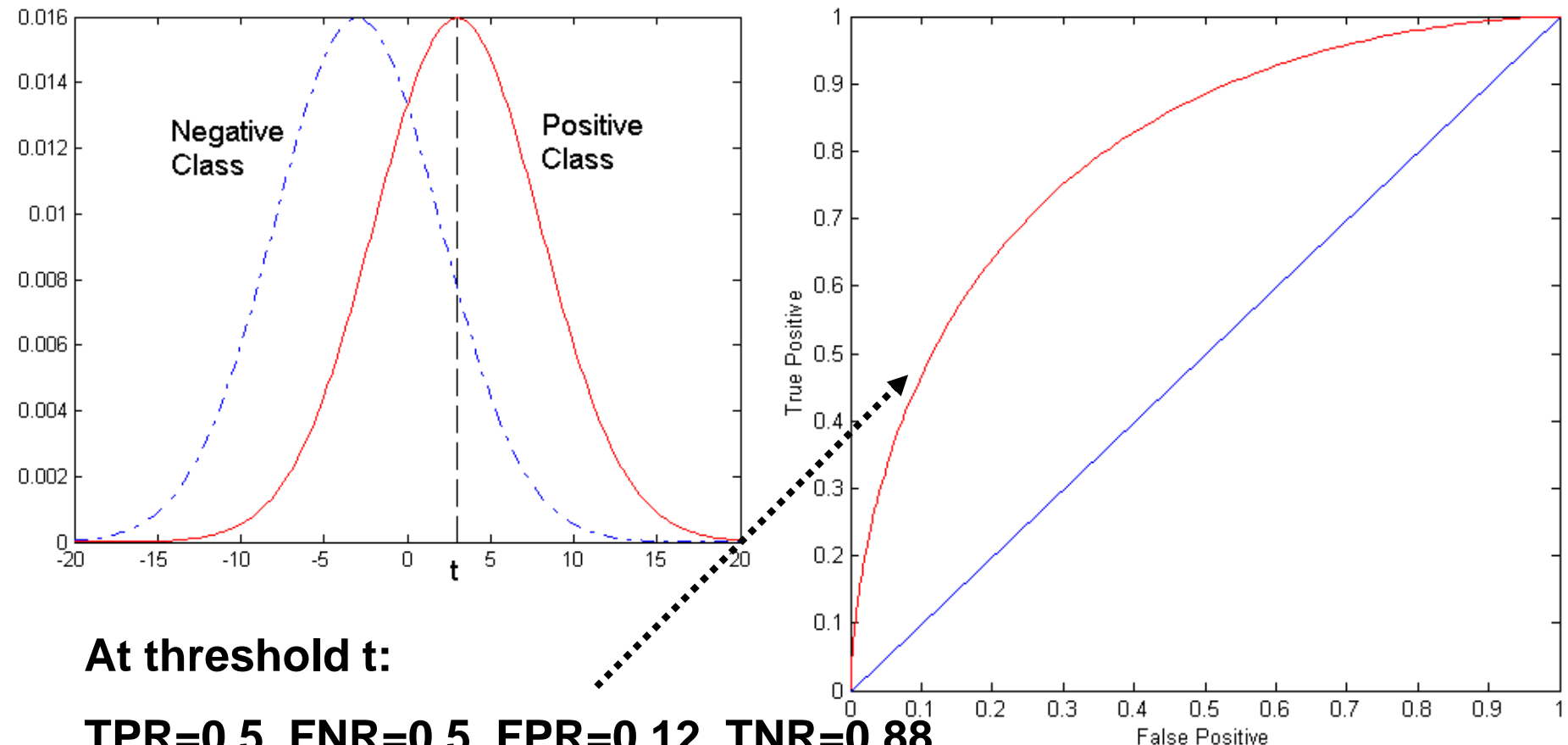
Using ROC for Model Comparison
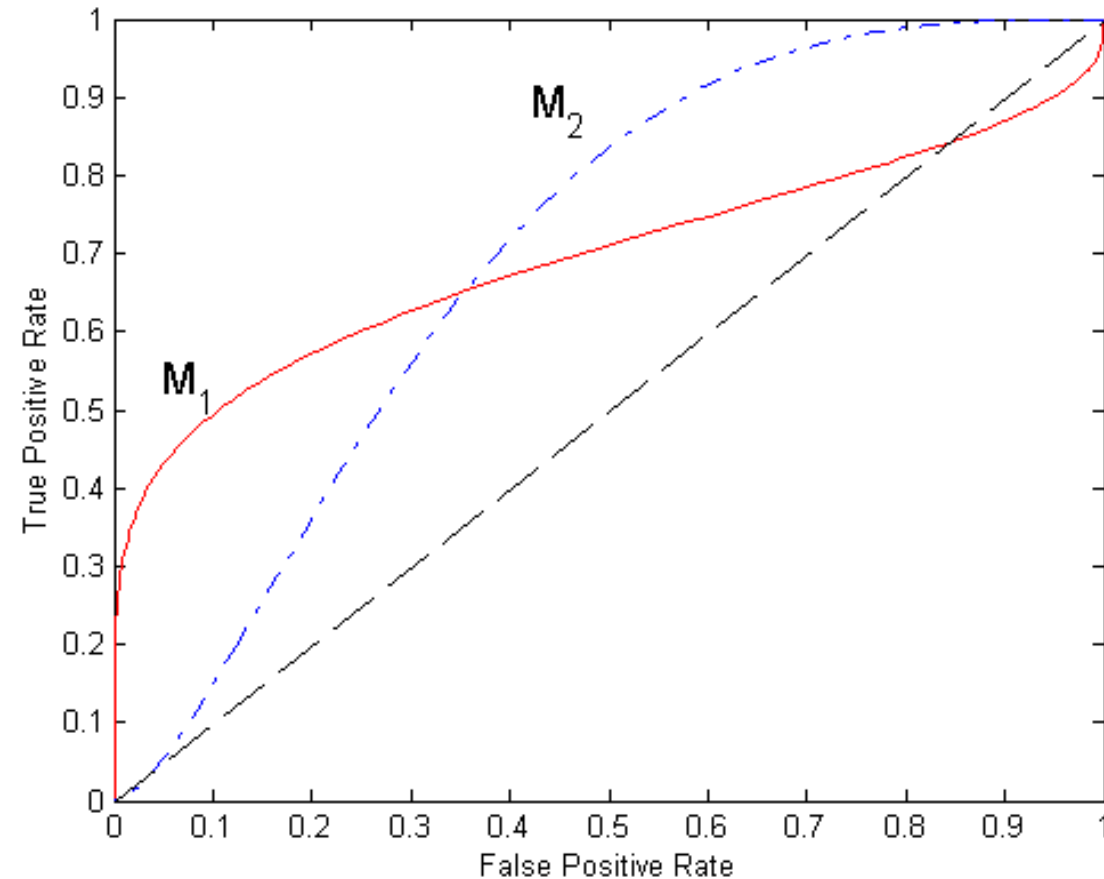
- No model consistently outperform the other
  - M_1 is better for small FPR
  - M_2 is better for large FPR

- Area Under the ROC curve
  - Ideal:
    - Area = 1
  - Random guess:
    - Area = 0.5

How to Construct an ROC curve

Instance	Score	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

- Use a classifier that produces a continuous-valued score for each instance
  - The more likely it is for the instance to be in the + class, the higher the score
- Sort the instances in decreasing order according to the score
- Apply a threshold at each unique value of the score
- Count the number of TP, FP, TN, FN at each threshold
  - TPR = TP/(TP+FN)
  - FPR = FP/(FP + TN)

How to construct an ROC curve

Class	+	-	+	-	-	-	+	-	+	+	
Threshold >=	0.25	0.43	0.53	0.76	0.85	0.85	0.85	0.87	0.93	0.95	1.00
TP	5	4	4	3	3	3	3	2	2	1	0
FP	5	5	4	4	3	2	1	1	0	0	0
TN	0	0	1	1	2	3	4	4	5	5	5
FN	0	1	1	2	2	2	2	3	3	4	5
TPR	1	0.8	0.8	0.6	0.6	0.6	0.6	0.4	0.4	0.2	0
FPR	1	1	0.8	0.8	0.6	0.4	0.2	0.2	0	0	0

ROC Curve: