CS722/822: Machine Learning

Instructor: Jiangwen Sun

Computer Science Department

Model Evaluation

- Metrics for Performance Evaluation
  - How to evaluate the performance of a model?

- Methods for Performance Evaluation
  - How to obtain reliable estimates?

- Methods for Model Comparison
  - How to compare the relative performance among competing models?

ROC (Receiver Operating Characteristic)

- Developed in 1950s for signal detection theory to analyze noisy signals
  - Characterize the trade-off between positive hits and false alarms
- ROC curve plots TPR (on the y-axis) against FPR (on the x-axis)
- A pair of (TPR, FPR) is represented as a point on the ROC curve
- Adjusting certain parameter of a detector can generate a series of (TPR, FPR) pairs, which lead to a curve
- A classifier has similar behavior to a signal detector, when it returns a real-valued prediction
  - Also need to balance the TPR and FPR
  - Changing the classification threshold changes the location of the point

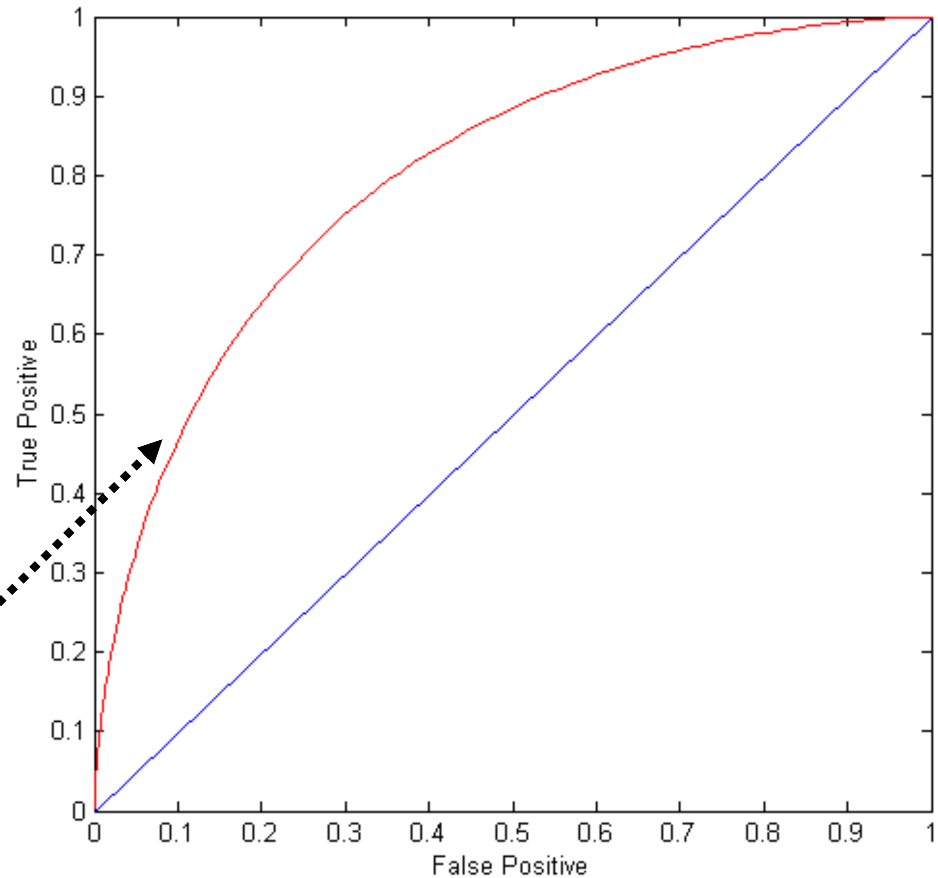
ROC Curve

	PREDICTED CLASS		
		Class =Yes	Class= No
ACTUAL CLASS	Class =Yes	a (TP)	b (FN)
	Class =No	c (FP)	d (TN)

TPR = TP/(TP+FN)
FPR = FP/(FP+TN)

At threshold t:

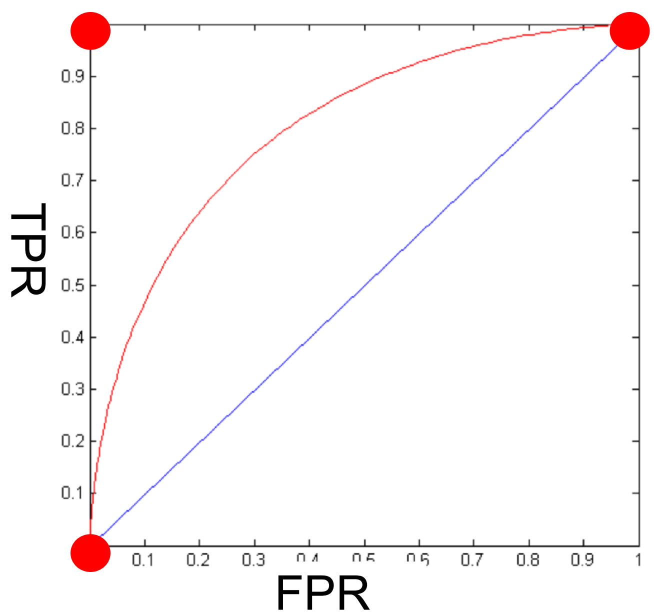
TP=50, FN=50, FP=12, TN=88



ROC Curve

		PREDICTED CLASS	
		Class=Yes	Class=No
ACTUAL CLASS	Class=Yes	a (TP)	b (FN)
	Class=No	c (FP)	d (TN)

TPR = TP/(TP+FN); **FPR** = FP/(FP+TN)



(TPR,FPR):

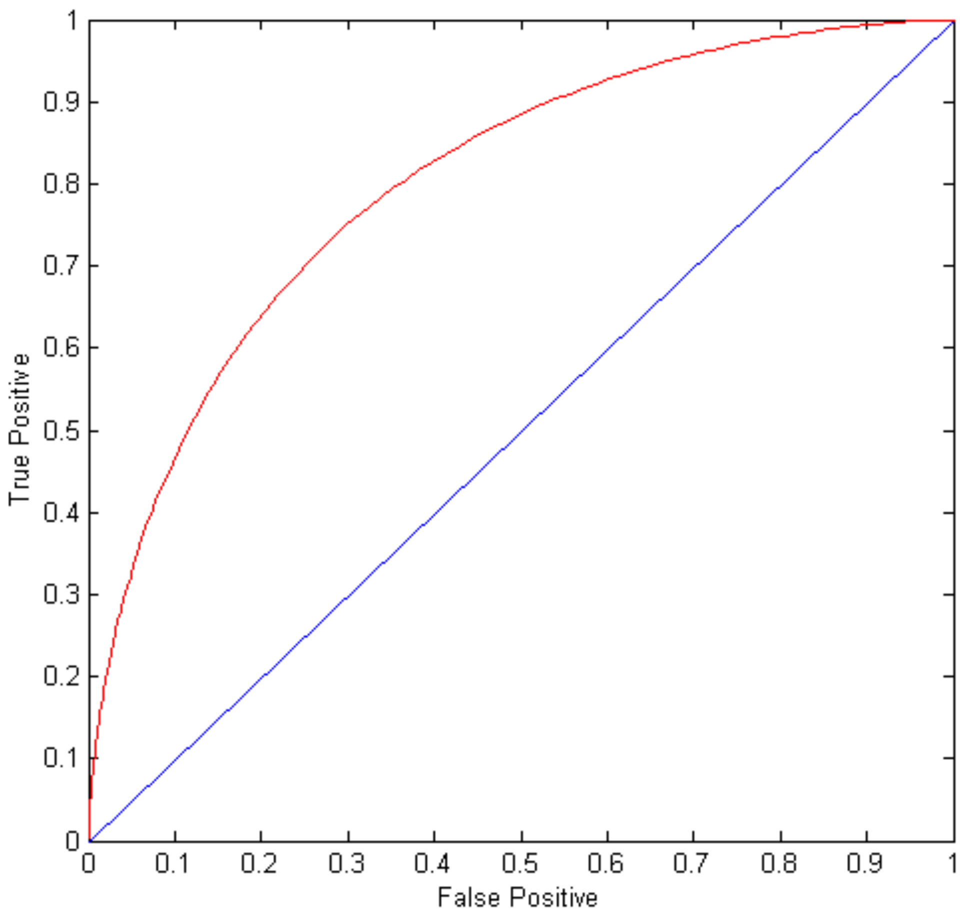
- (0,0): declare everything to be negative class
  - TP=0, FP = 0
- (1,1): declare everything to be positive class
  - FN = 0, TN = 0
- (1,0): ideal
  - FN = 0, FP = 0

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



How to Construct an ROC curve

Instance	P(+|A)	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 classifier that produces posterior probability for each test instance P(+|A)
- Sort the instances according to P(+|A) in decreasing order
- Apply threshold at each unique value of P(+|A)
- Count the number of TP, FP, TN, FN at each threshold
- TP rate, TPR = TP/(TP+FN)
- FP rate, FPR = FP/(FP + TN)

How to Construct an ROC curve

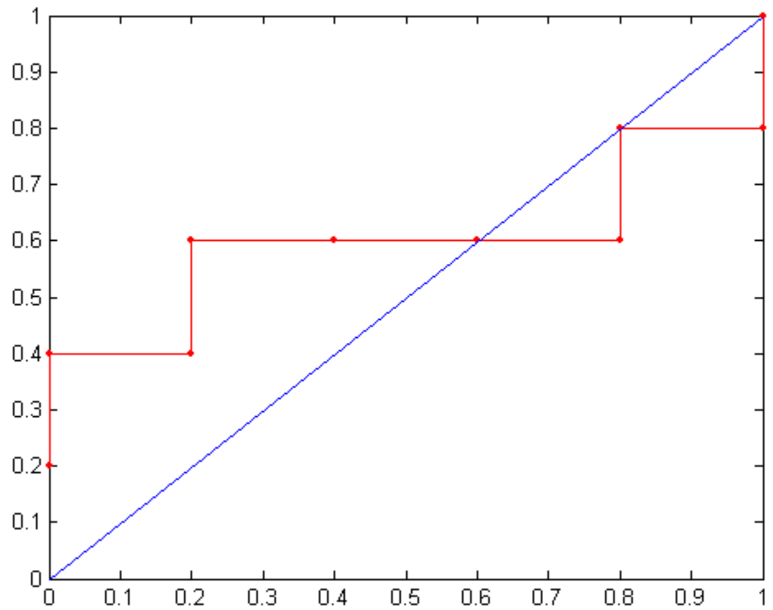
Instance	P(+|A)	True Class
1	0.95	+
2	0.93	+
3	0.87	-
4	0.85	-
5	0.85	-
6	0.85	+
7	0.76	-
8	0.53	+
9	0.43	-
10	0.25	+

- Use classifier that produces posterior probability for each test instance P(+|A)
- Sort the instances according to P(+|A) in decreasing order
- **Pick a threshold 0.85**
- p>= 0.85, predicted to P
- p< 0.85, predicted to N
- TP = 3, FP=3, TN=2, FN=2
- TP rate, TPR = 3/5=60%
- FP rate, FPR = 3/5=60%

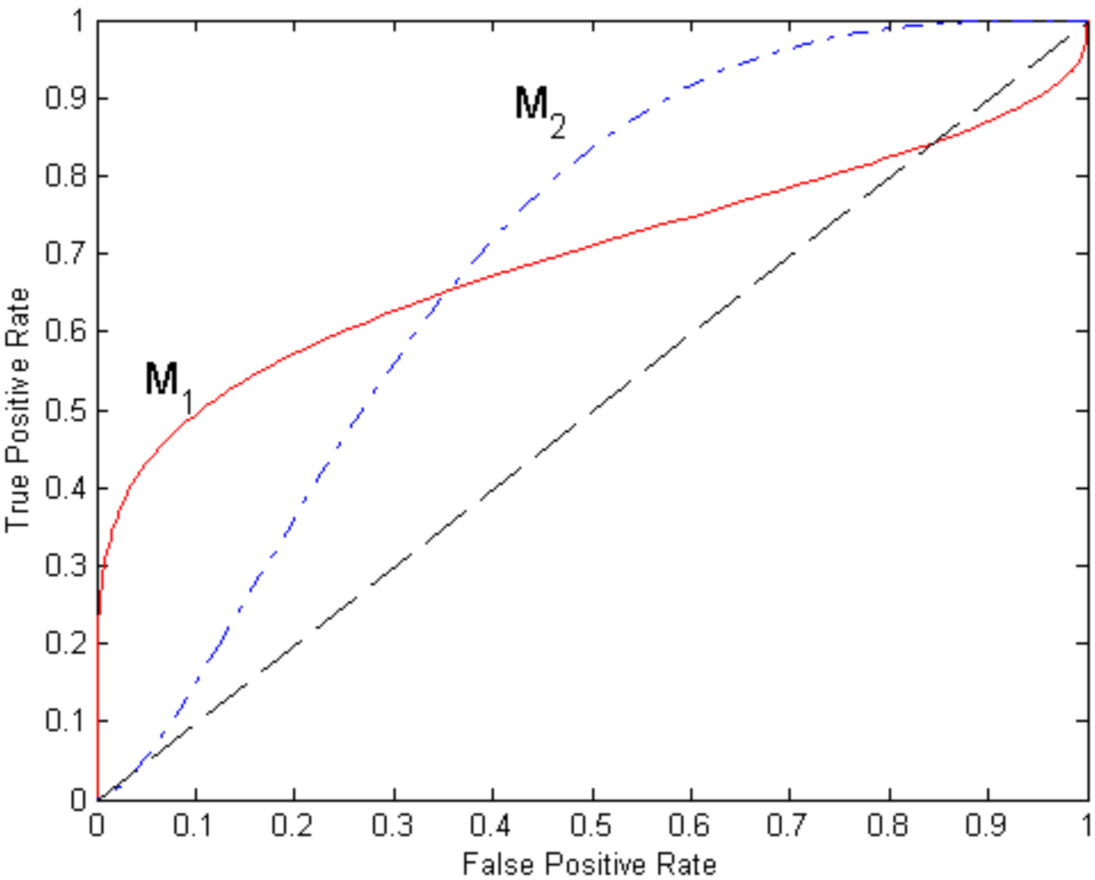
How to construct an ROC curve

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:



Using ROC for Model Comparison

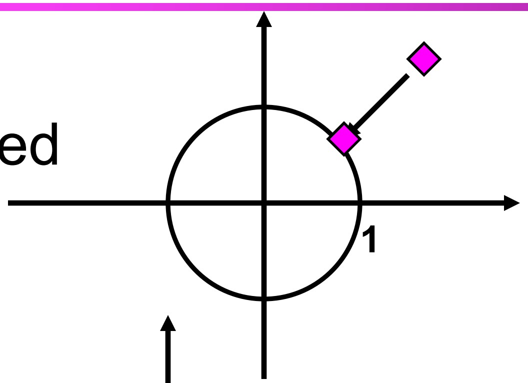


- No model consistently outperforms the other
  - M_1 is better for small FPR
  - M_2 is better for large FPR
- Area Under the ROC curve (AUC)
  - Ideal:
    - Area = 1
  - Random guess:
    - Area = 0.5

Data normalization

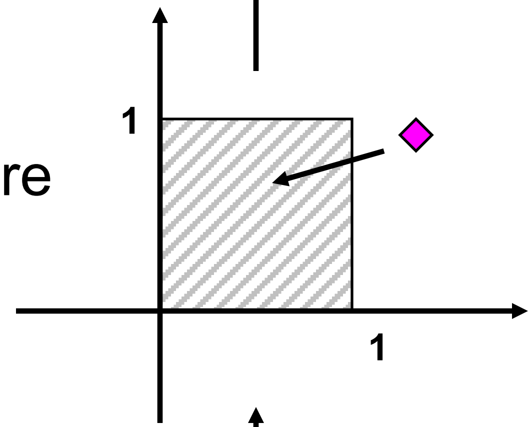
- Example-wise normalization
  - Each example i is normalized and mapped to unit sphere

$$\mathbf{x}^{(i)} / \|\mathbf{x}^{(i)}\|$$



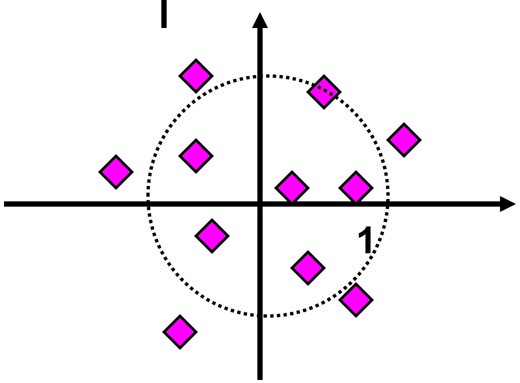
- Feature-wise normalization
  - [0,1]-normalization: normalize each feature i into a unit space

$$(\mathbf{x}_i - \min(\mathbf{x}_i)) / (\max(\mathbf{x}_i) - \min(\mathbf{x}_i))$$



$\mathbf{x}_i$ is the data vector of feature i, $\min(\mathbf{x}_i)$ is the minimum value in the vector, $\max(\mathbf{x}_i)$ is the maximum value

  - Standard normalization: normalize each feature i to have mean 0 and standard deviation 1



$$(\mathbf{x}_i - \mu_i) / \sigma_i$$

μ_i and σ_i is the sample mean and variance of feature i