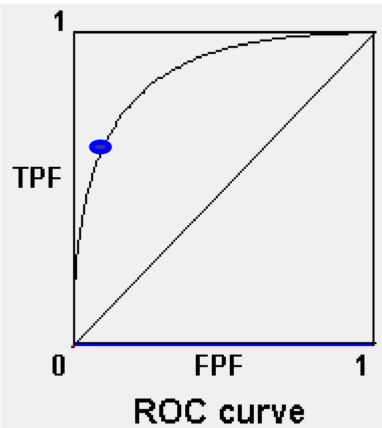
ROC Curves

Lecture 22

by Marina Barsky

ROC (Receiver Operating Characteristic)



True positive and False positive fractions are plotted as we move the dividing threshold.

Tradeoff between a true signal rate and the false alarm rate

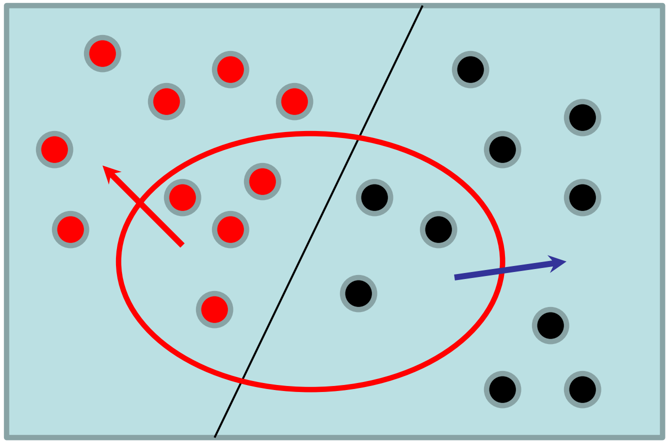
- Part of "Signal Detection Theory" developed during World War II for the analysis of radar images: Radar operators had to decide whether a blip on the screen represented an enemy target, a friendly ship, or just a noise
- Signal detection theory measures the ability of radar receivers to make these important distinctions
- Their ability to do so was called the

__R__eceiver __O__perating __C__haracteristics (__ROC__)

True Positive Rate (Fraction) and False Positive Rate (Fraction)

		Predicted class		
		+	-	Total
Actual class	+	True positive	False negative	Positives
	-	False positive	True negative	Negatives

Confusion matrix

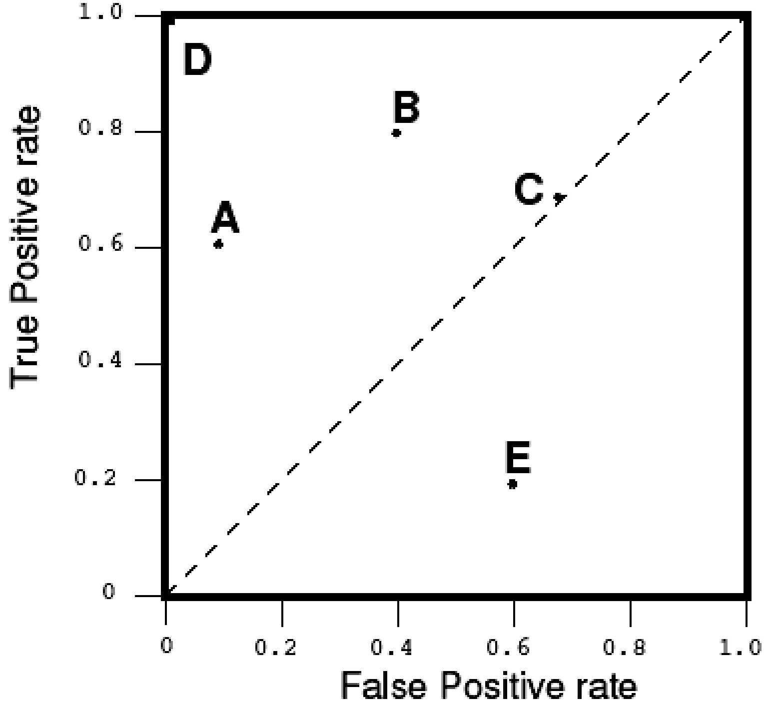


TPF= P correctly classified as P / P

FPF= N incorrectly classified as P / N

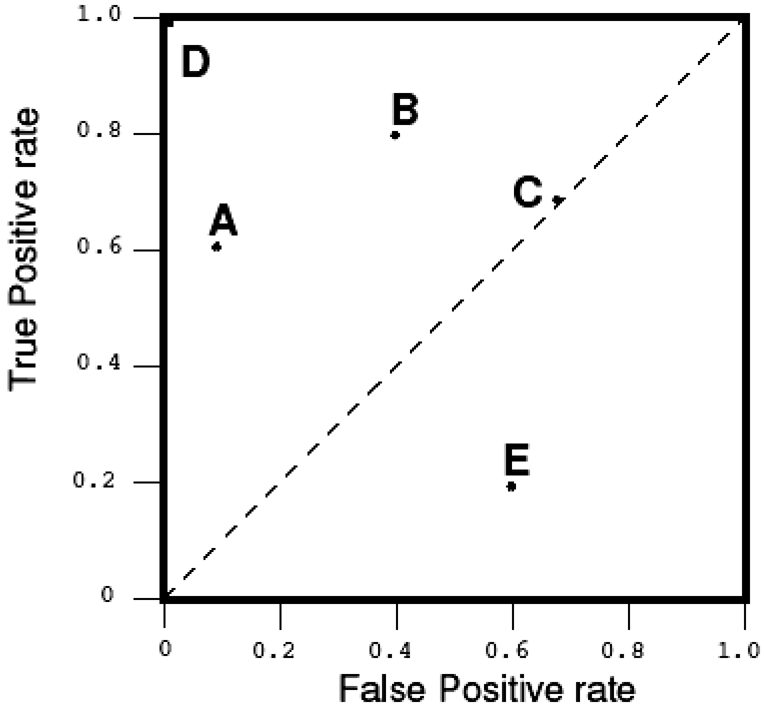
ROC Space

- ROC graphs are two-dimensional graphs in which TP rate is plotted on the Y axis and FP rate is plotted on the X axis.
- ROC graph depicts relative trade-offs between benefits (true positives) and costs (false positives).



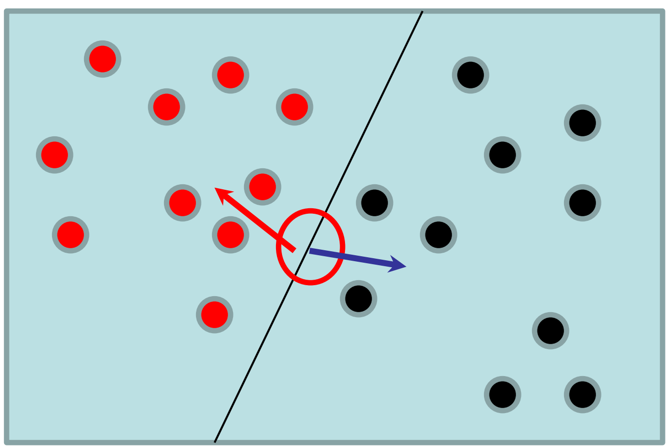
ROC Space: discrete classifiers

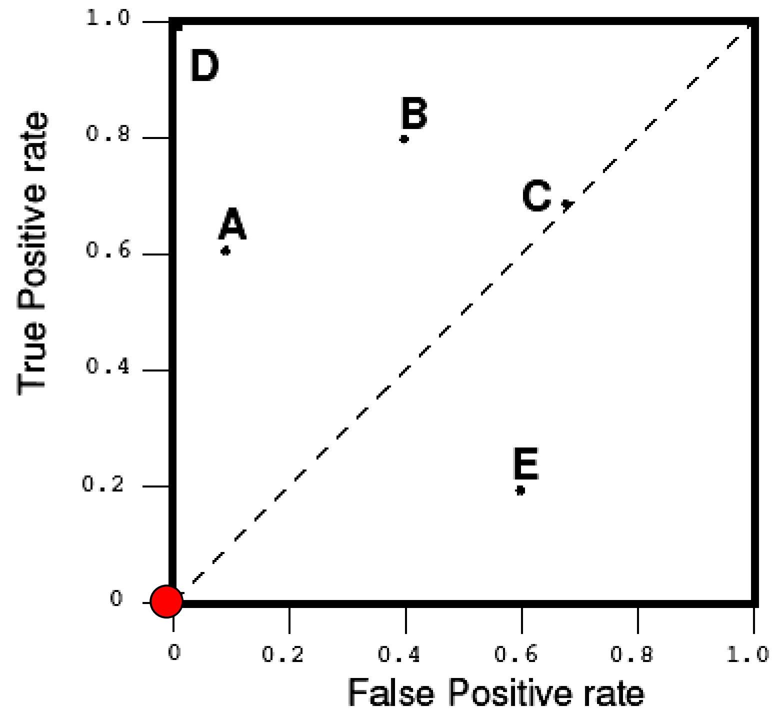
- Figure shows a ROC graph with five classifiers labeled A through E.
- A *discrete classier* is one that outputs only a class label.
- Each discrete classier produces a confusion matrix. We take a pair (fp rate, tp rate)and plot it as a single point in ROC space.
- Classifiers in a figure are all discrete classifiers.



Special Points in ROC Space

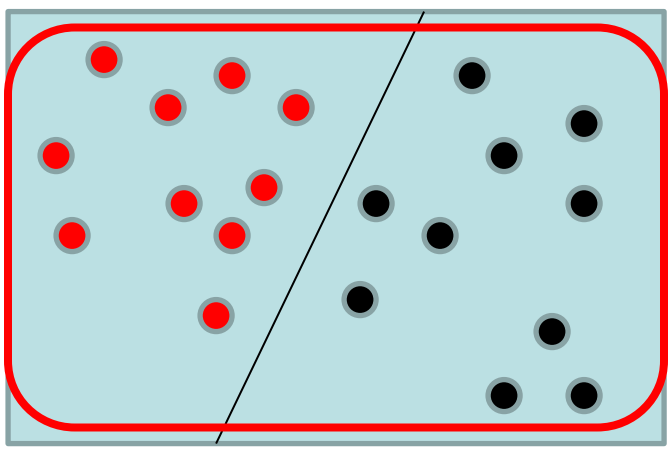
- Lower left point (0, 0) represents the strategy of never issuing a positive classification
  - such a classier commits no false positive errors
  - but also gains no true positives

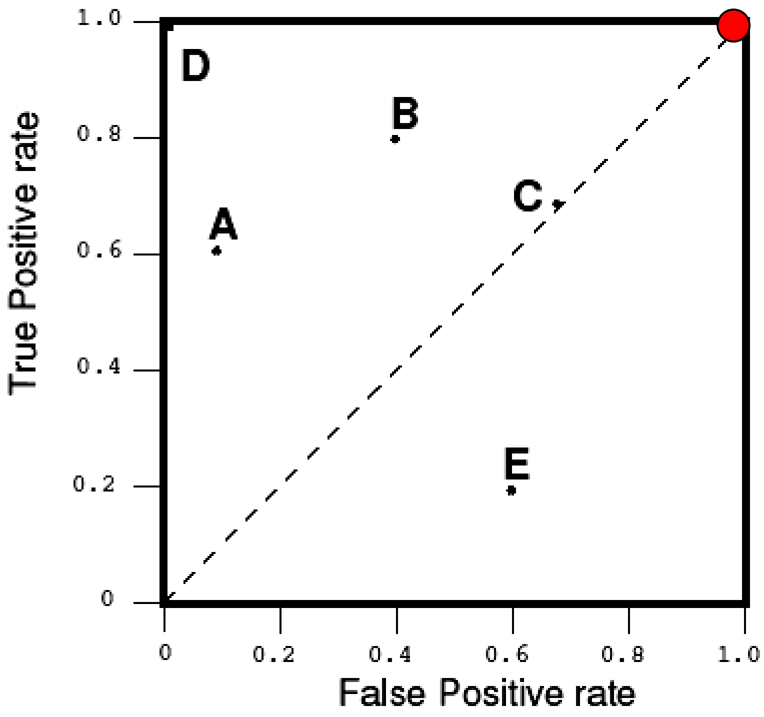




Special Points in ROC Space

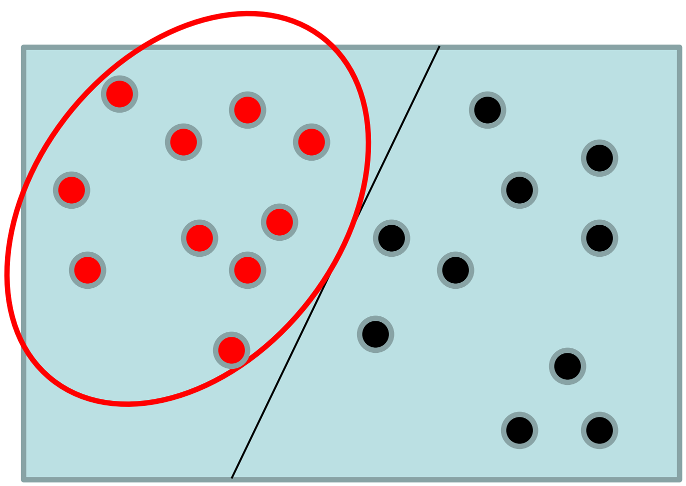
- Upper right corner (1, 1) represents the opposite strategy, of unconditionally issuing positive classifications

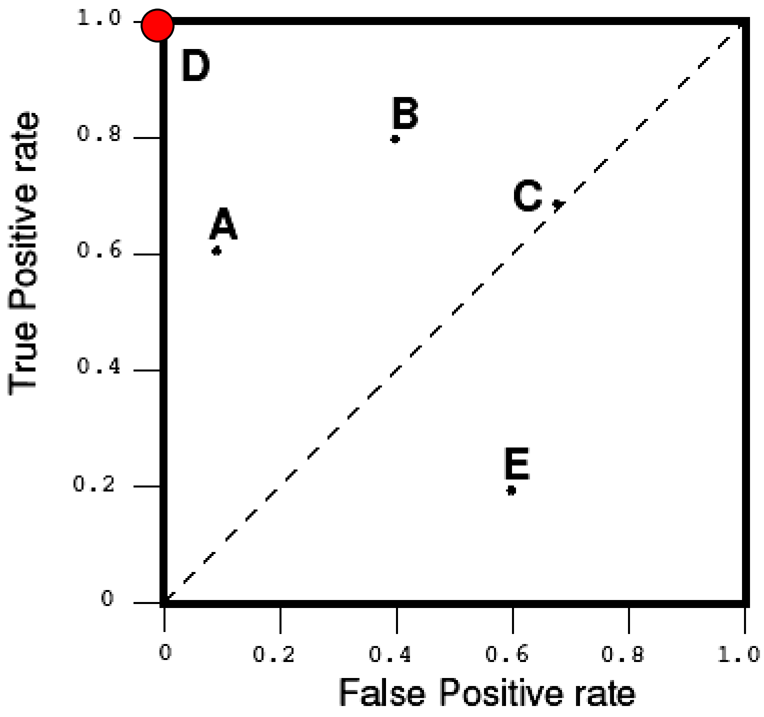




Special Points in ROC Space

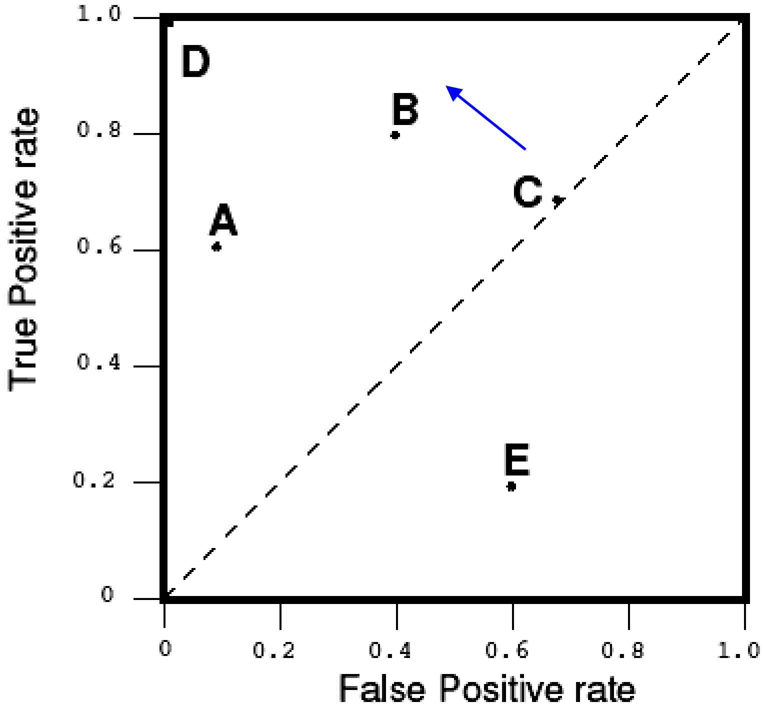
- Point (0, 1) represents perfect classification.
  - D's performance is perfect as shown





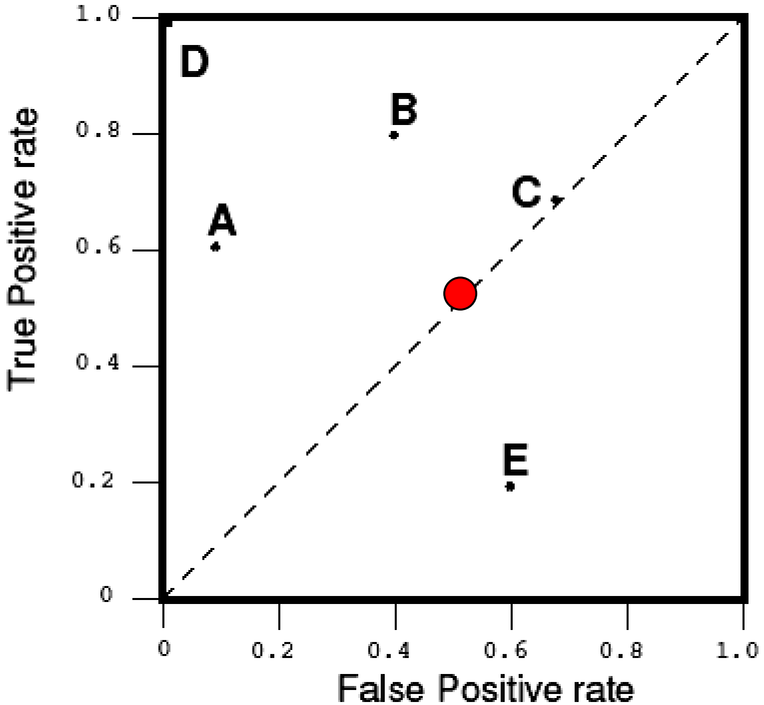
A point in ROC space

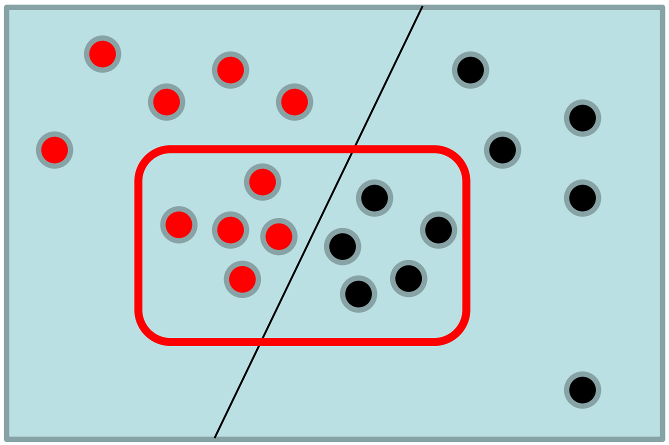
- Informally, one point in ROC space is better than another if it is to the northwest of the first
  - tp rate is higher, fp rate is lower, or both



Random Classifiers

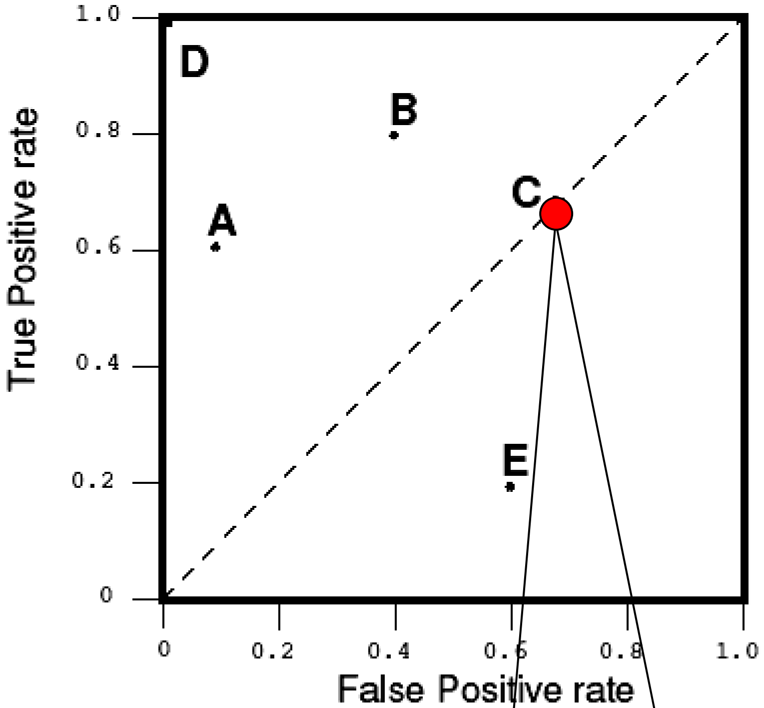
- The diagonal line y = x represents the strategy of randomly guessing a class.
- For example, if a classier randomly says "Positive" half the time (regardless of the instance provided), it can be expected to get half the positives and half the negatives correctly:
  - this yields the point (0.5, 0.5)



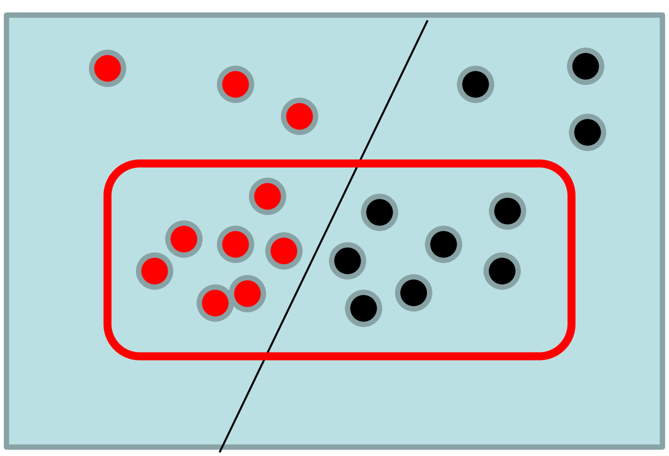


Random Classifiers

- If it randomly says “Positive” 70% of the time (regardless of the instance provided), it can be expected to:
  - get 70% of the positives correct, but
  - its false positive rate will increase to 70% as well, yielding (0.7, 0.7) in ROC space

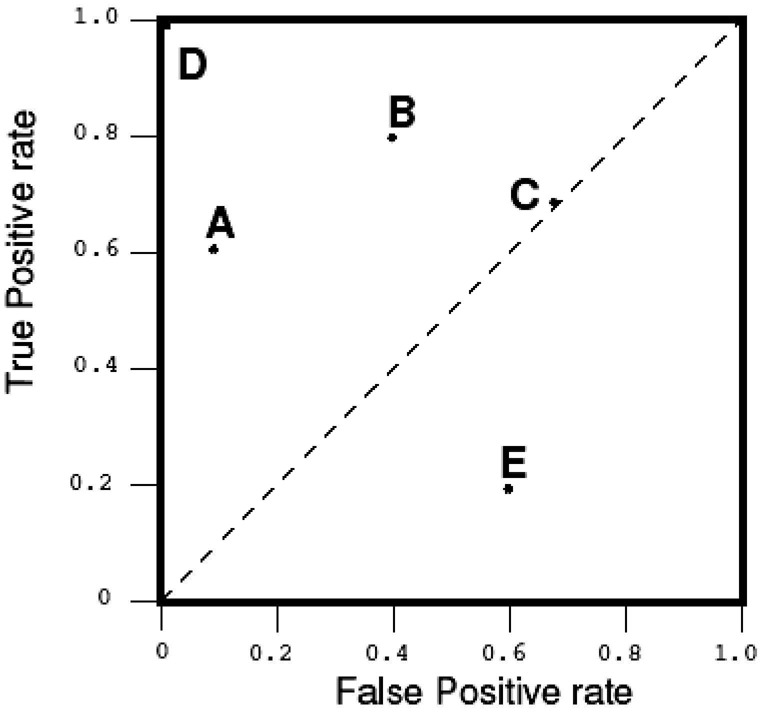


C's performance is virtually random.
At (0.7; 0.7), C is guessing the positive class 70% of the time.



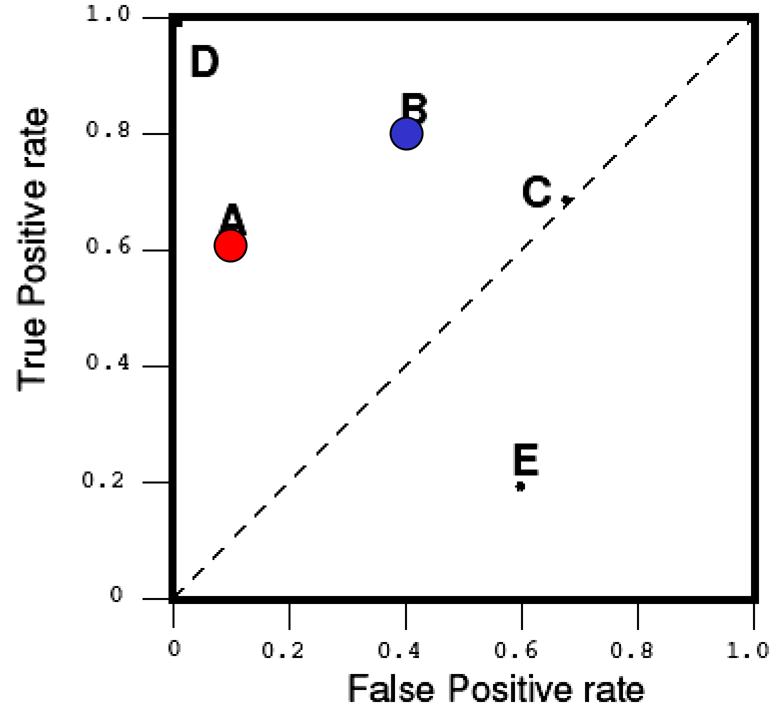
Random Classifiers

- A random classier will produce a ROC point that "slides" back and forth on the diagonal based on the frequency with which it ‘guesses’ the positive class



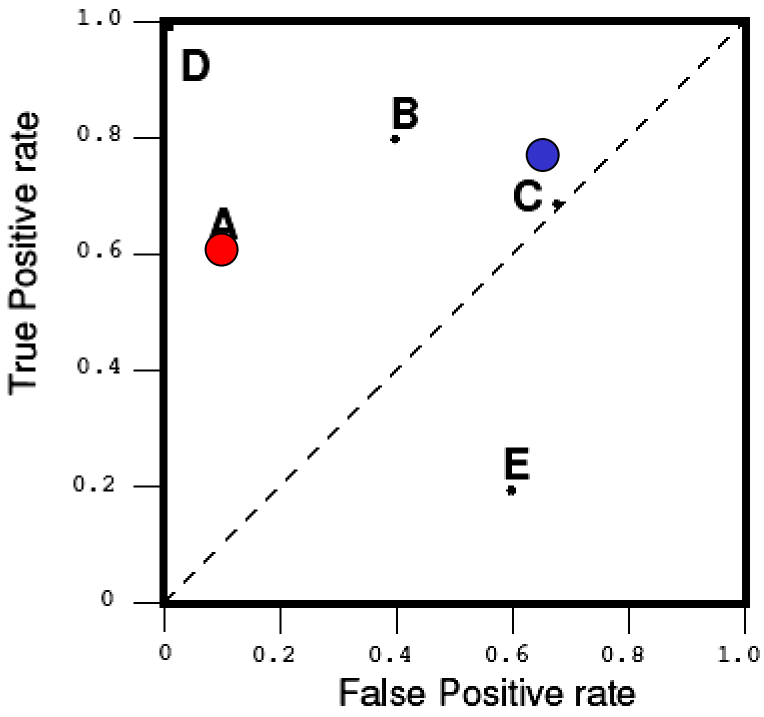
"Conservative" vs. "Liberal"

- Classifiers appearing on the left hand-side of an ROC graph, near the Y axis, may be thought of as "conservative"
  - they make positive classifications only with strong evidence so they make few false positive errors
  - but may have low true positive rates as well
- In figure, A is more conservative than B.



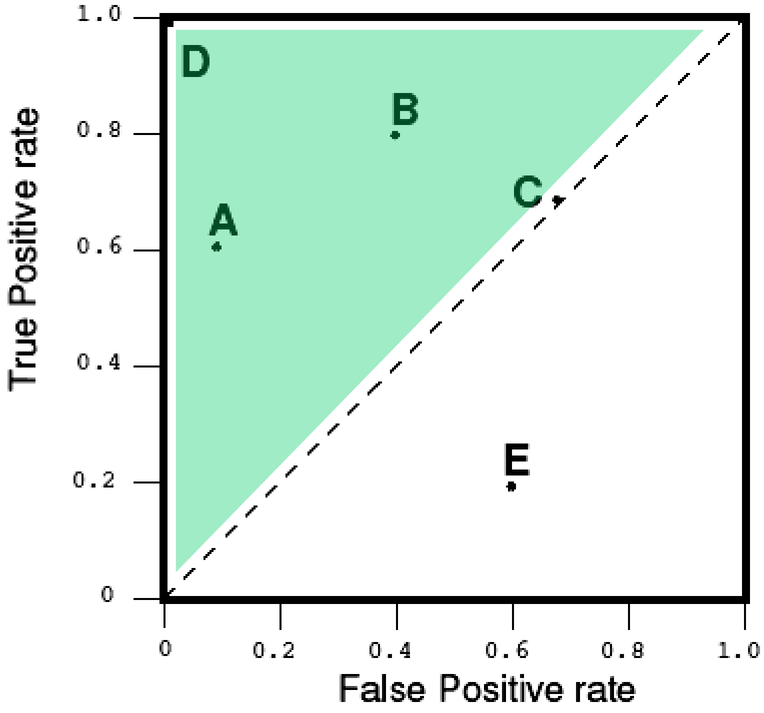
“Conservative” vs. “Liberal”

- Classifiers on the upper right-hand side of an ROC graph may be thought of as “liberal”
  - they make positive classifications with weak evidence so they classify nearly all positives correctly
  - but they may have high false positive rates
- In figure, C is more liberal than A.



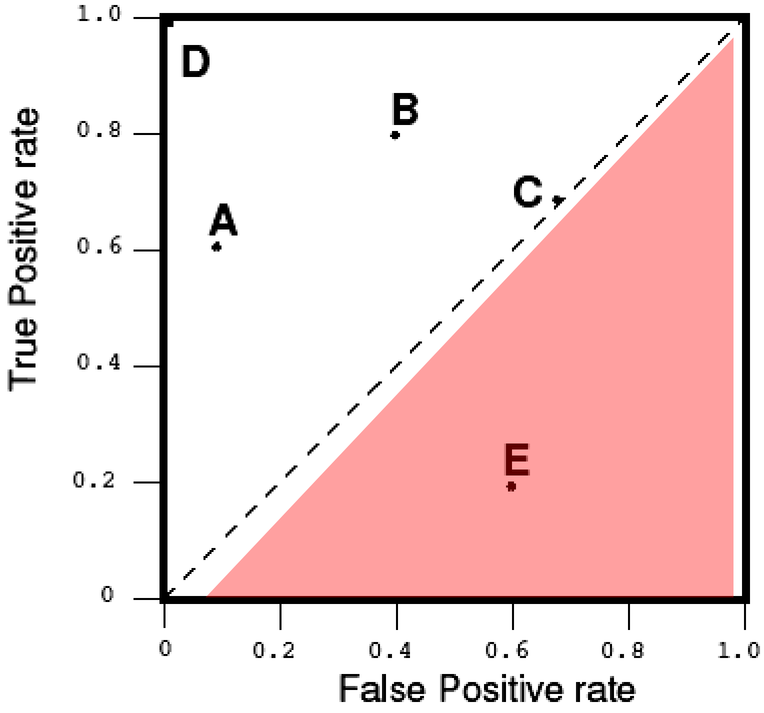
Non-random classifier: Upper Triangular Area

- To get away from the diagonal into the upper triangular region, the classifier must exploit some information in the data



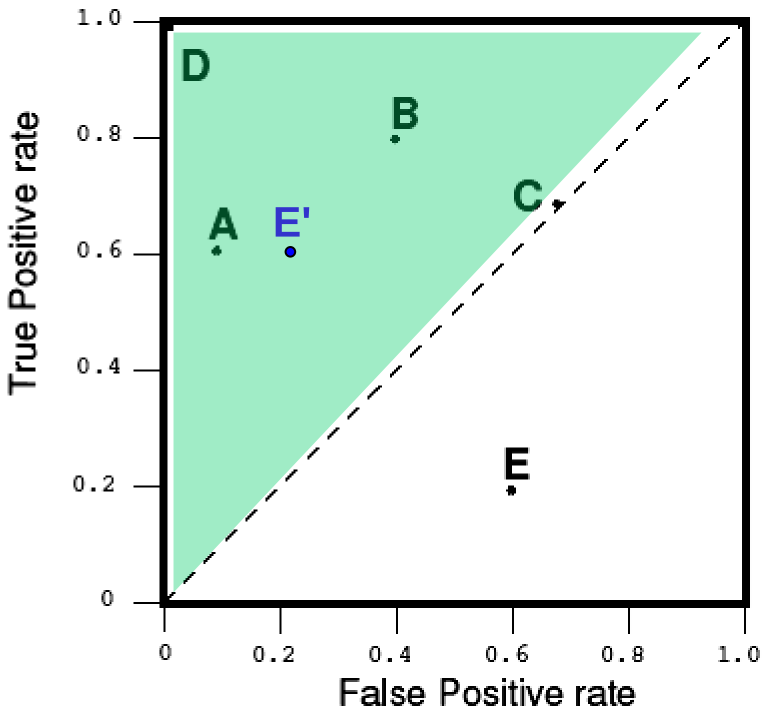
Worse-than-random classifier: Lower Triangular Area

- Any classifier that appears in the lower right triangle performs worse than random guessing
  - This triangle is therefore usually empty in ROC graphs



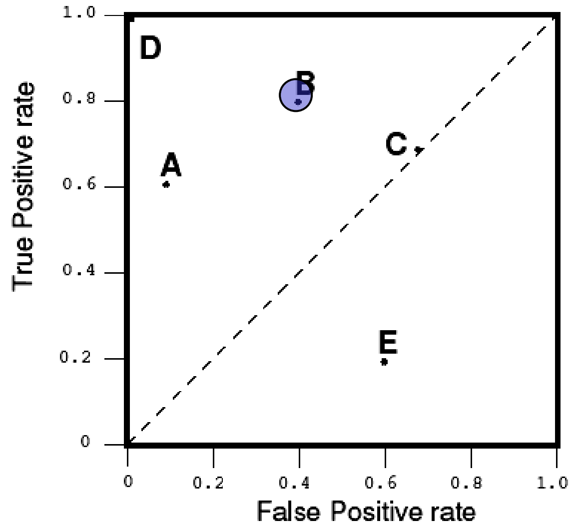
Worse-than-random classifier may be useful

- If we negate a classifier - that is, reverse its classification decisions on every instance, then:
  - its true positive classifications become false negative mistakes, and
  - its false positives become true negatives.
- A classifier below the diagonal may be said to have discovered a useful information, but it is applying the information incorrectly



Curves and points in ROC space

- Many classifiers, such as decision trees or rule learners, are designed to produce only a class decision, i.e., a Y or N on each instance.
  - When such a discrete classier is applied to a test set, it yields a single confusion matrix, which in turn corresponds to one ROC point
  - Thus, a discrete classifier produces only a single point in ROC space

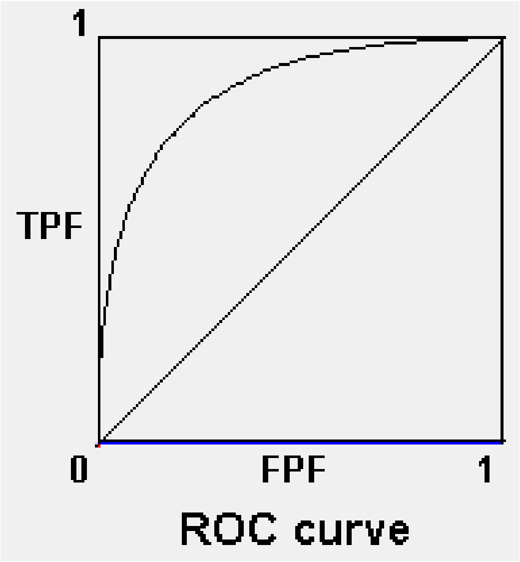


		Predicted class	
		Class +	Class -
Actual class	Class +	80	20
	Class -	40	60

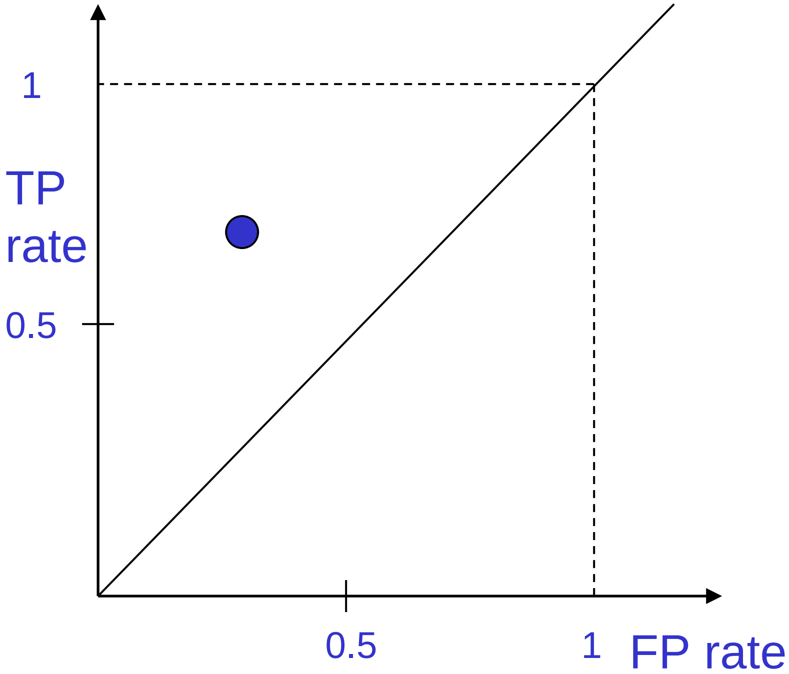
Confusion matrix for classifier B

Curves and points in ROC space

- Probabilistic classifiers output a range of probabilities for each classified instance
  - We can generate a *ROC **curve*** for a probabilistic classifier



50% threshold



It is common to give the 50% threshold for positive classification

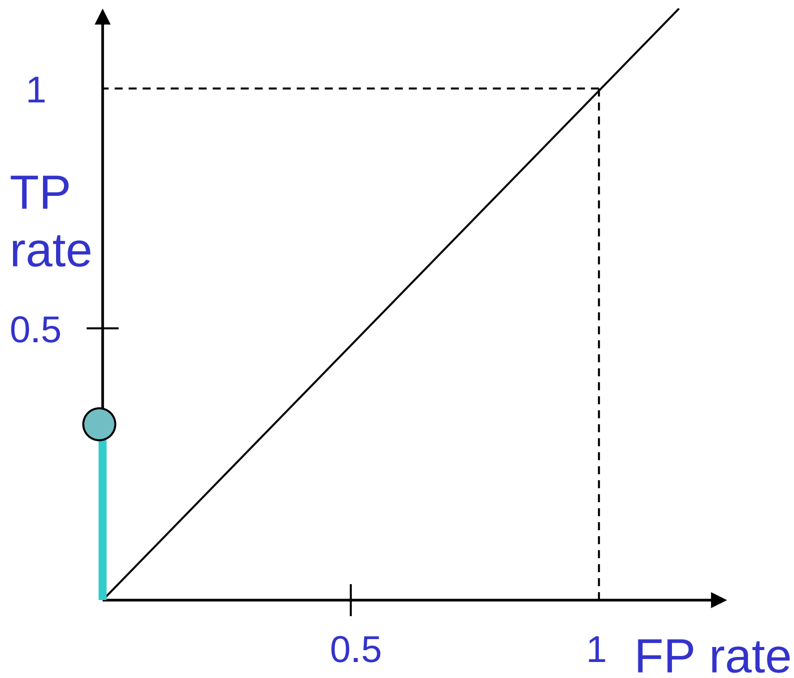
Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	NO	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

FP rate: FP/N=1/4=0.25

TP rate: TP/P=2/3≈0.7

However, for different problems we can set **different thresholds**

ROC curve of a probabilistic classifier



For different threshold values we get different points in the ROC space

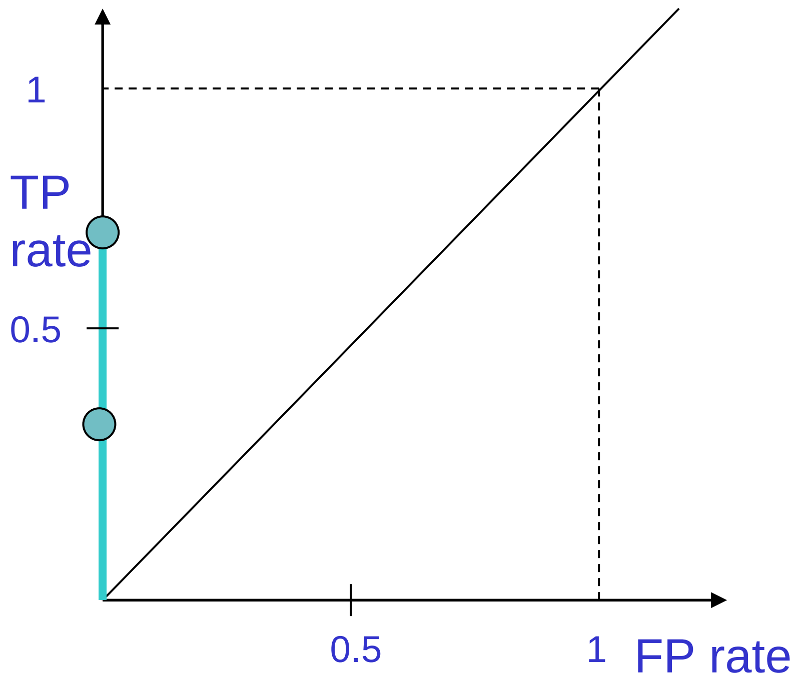
Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	NO	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

Operating threshold

FP rate: FP/N=0/4=0

TP rate: TP/P=1/3≈0.3

ROC curve of a probabilistic classifier



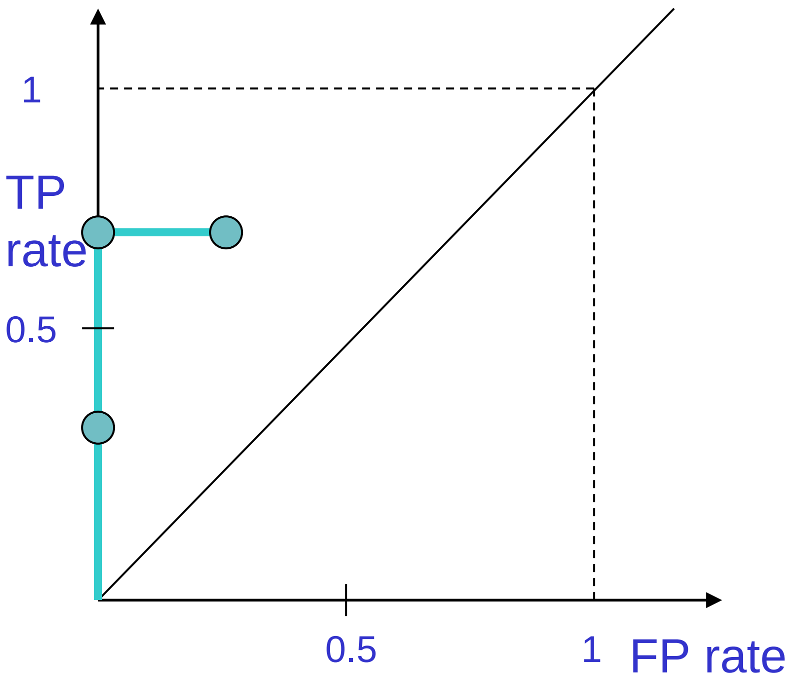
For different threshold values we get different points in the ROC space

Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	NO	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

FP rate: FP/N=0/4=0

TP rate: TP/P=2/3≈0.7

ROC curve of a probabilistic classifier



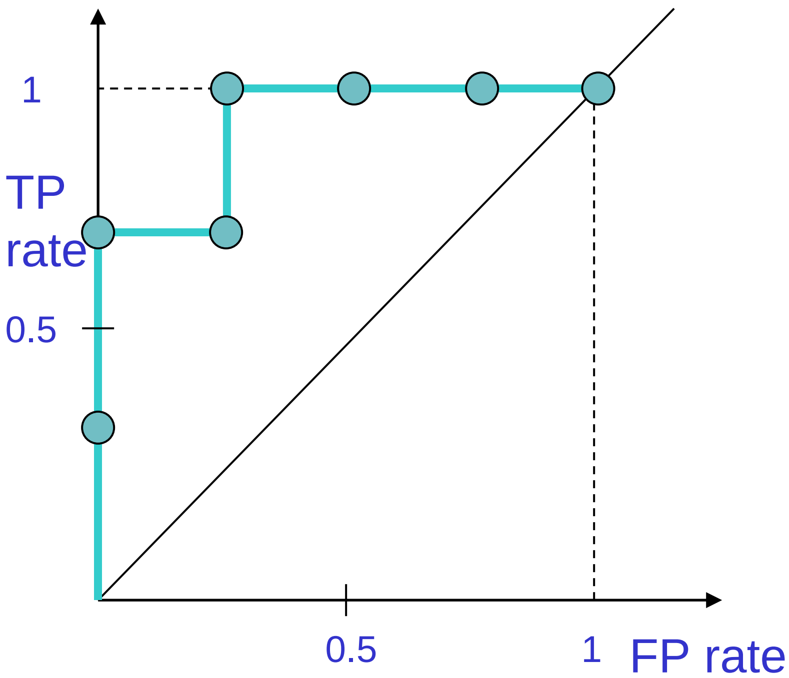
For different threshold values we get different points in the ROC space

Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	NO	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

FP rate: FP/N=1/4=0.25

TP rate: TP/P=2/3≈0.7

ROC curve of a probabilistic classifier



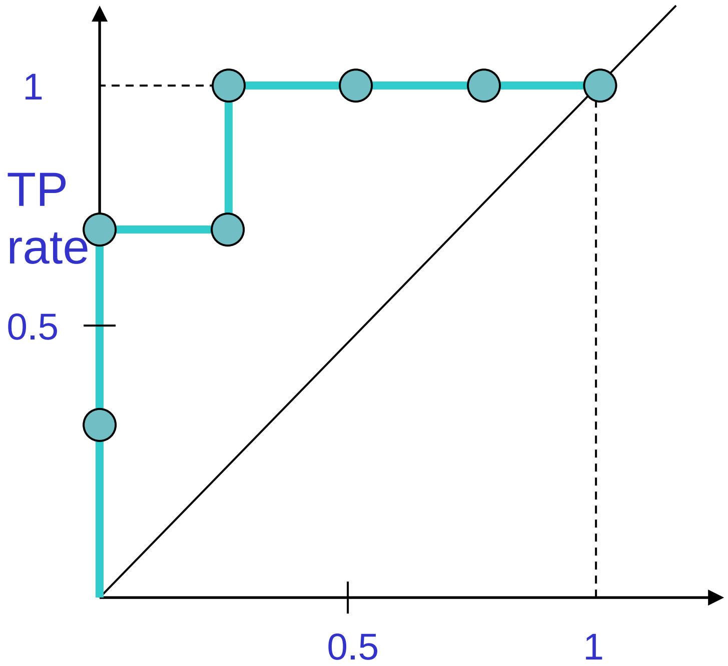
For different threshold values we get different points in the ROC space

Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	YES	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

FP rate: FP/N=1/4=0.25

TP rate: TP/P=3/3=1.0, etc…

ROC curve of a probabilistic classifier



At the end we get the ROC ***curve*** for a probabilistic classifier

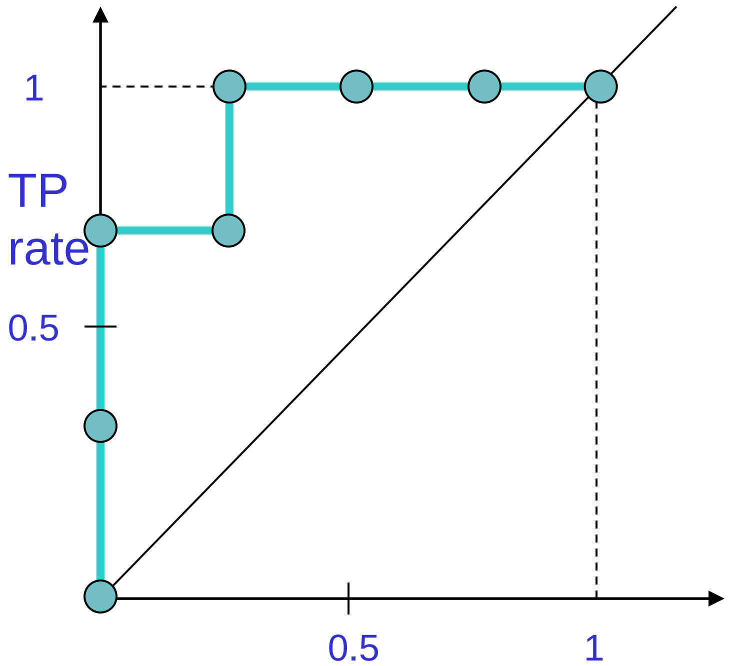
Building ROC curve: Algorithm

- Exploit monotonicity of thresholded classifications:
  - Any instance that is classified positive with respect to a given threshold will be classified positive for all lower thresholds as well.

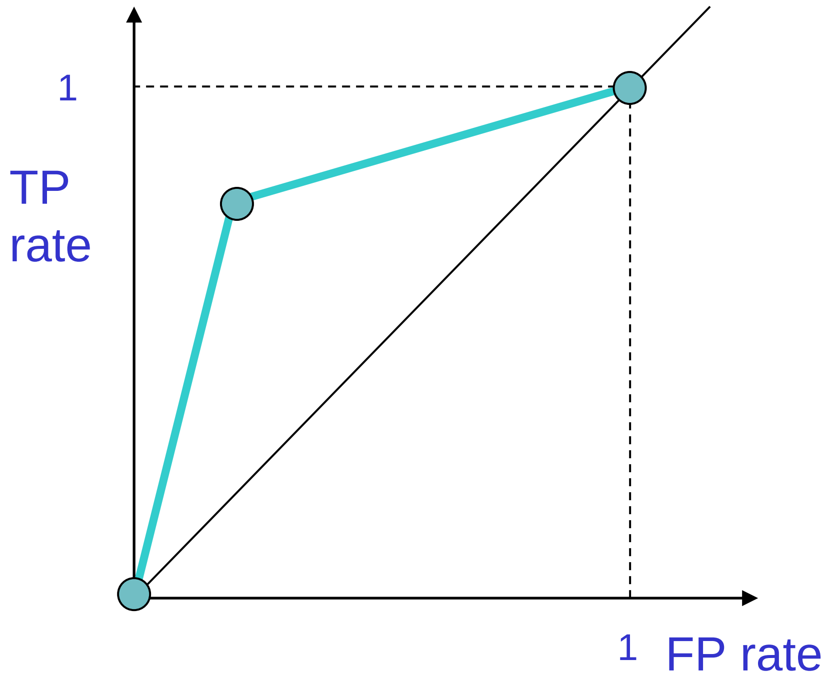
- Therefore, we can simply:
  - sort the test instances decreasing by their scores and
  - move down the list, processing one instance at a time and
  - updating TP and FP as we go.

- In this way, the ROC graph can be created from one linear scan.

ROC curve of a probabilistic classifier vs discrete classifier



ROC curve for Naïve Bayes classifier (probabilistic)



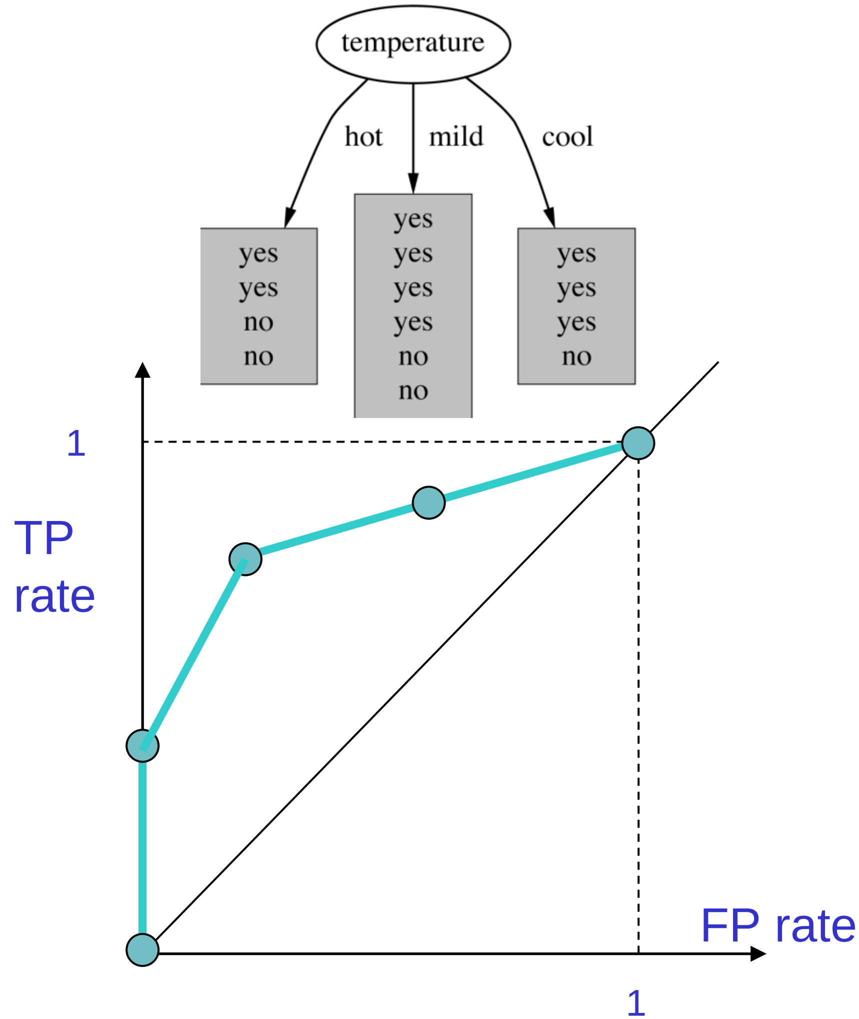
ROC *'curve'* for Decision Tree classifier (discrete)

ROC curve: of a probabilistic classifier vs discrete classifier

We can always convert a discrete classifier into a probabilistic classifier
For example, if we label each leaf of a pruned decision tree with the majority class, we can consider each positive prediction as a probability:

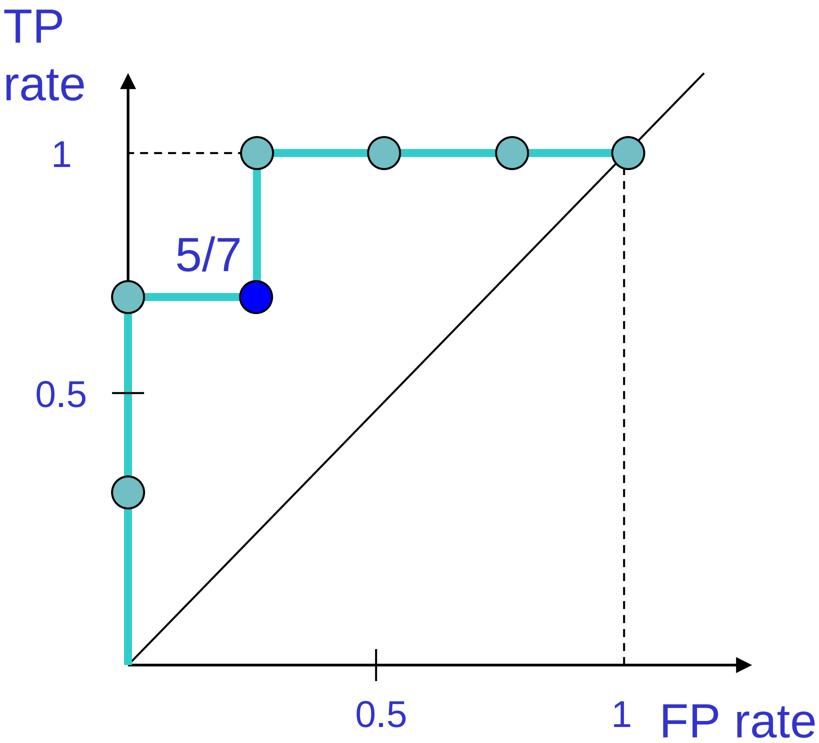
The leaf with 40 positives and 20 negatives is labeled as positive: +(P=40/60)

Then we can play with the operating threshold to create a real ROC curve



ROC ‘curve’ for Decision Tree classifier (probabilistic)

ROC curve: accuracy



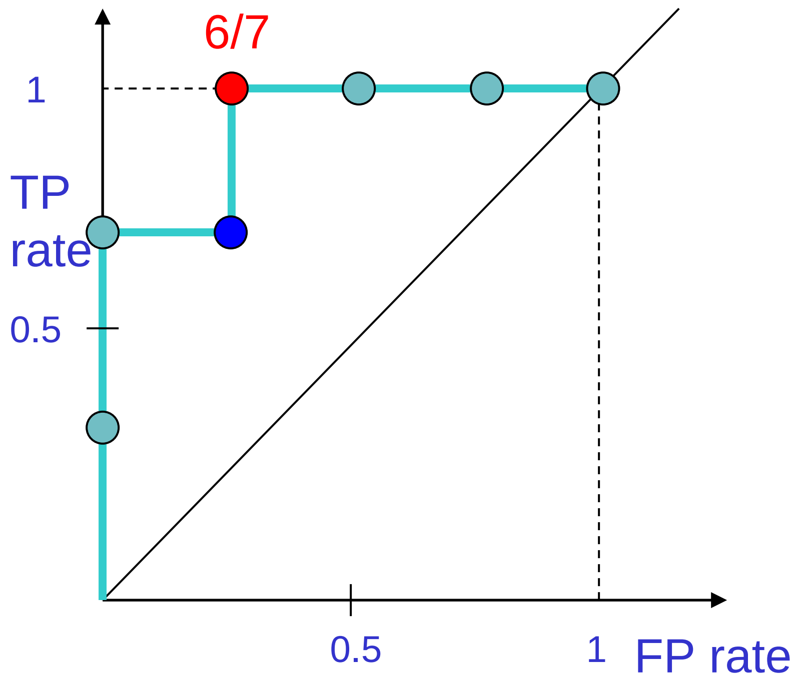
How many are correctly classified:

TPR*Pos+(1-FPR)*Neg=2+3=5

Accuracy: 5/7

Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	NO	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

ROC curve: accuracy (success rate)



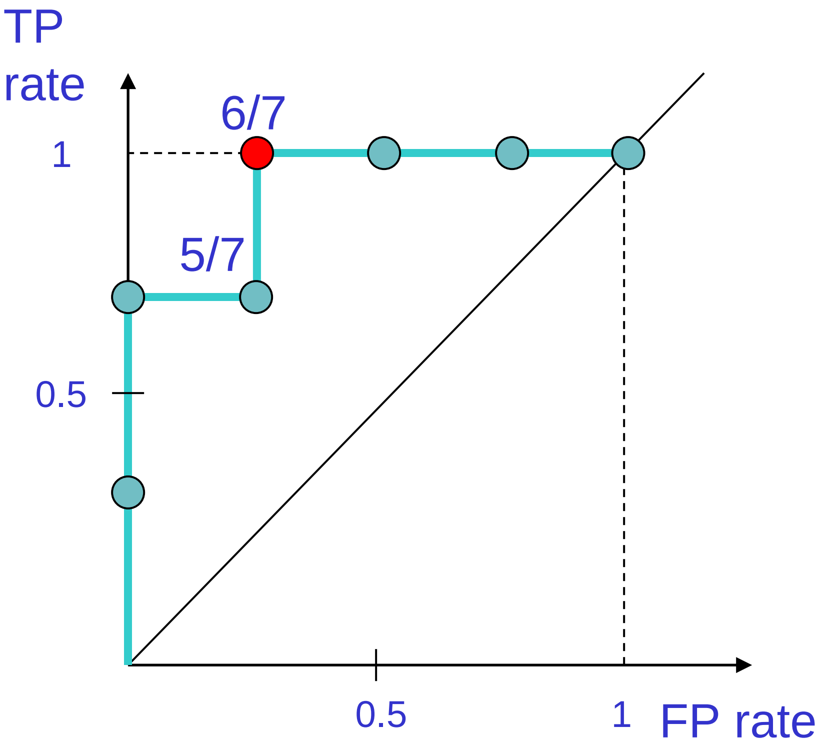
How many are correctly classified:

TPR*Pos+(1-FPR)*Neg=3+3=6

Accuracy: 6/7

Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	YES	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

ROC curve: accuracy



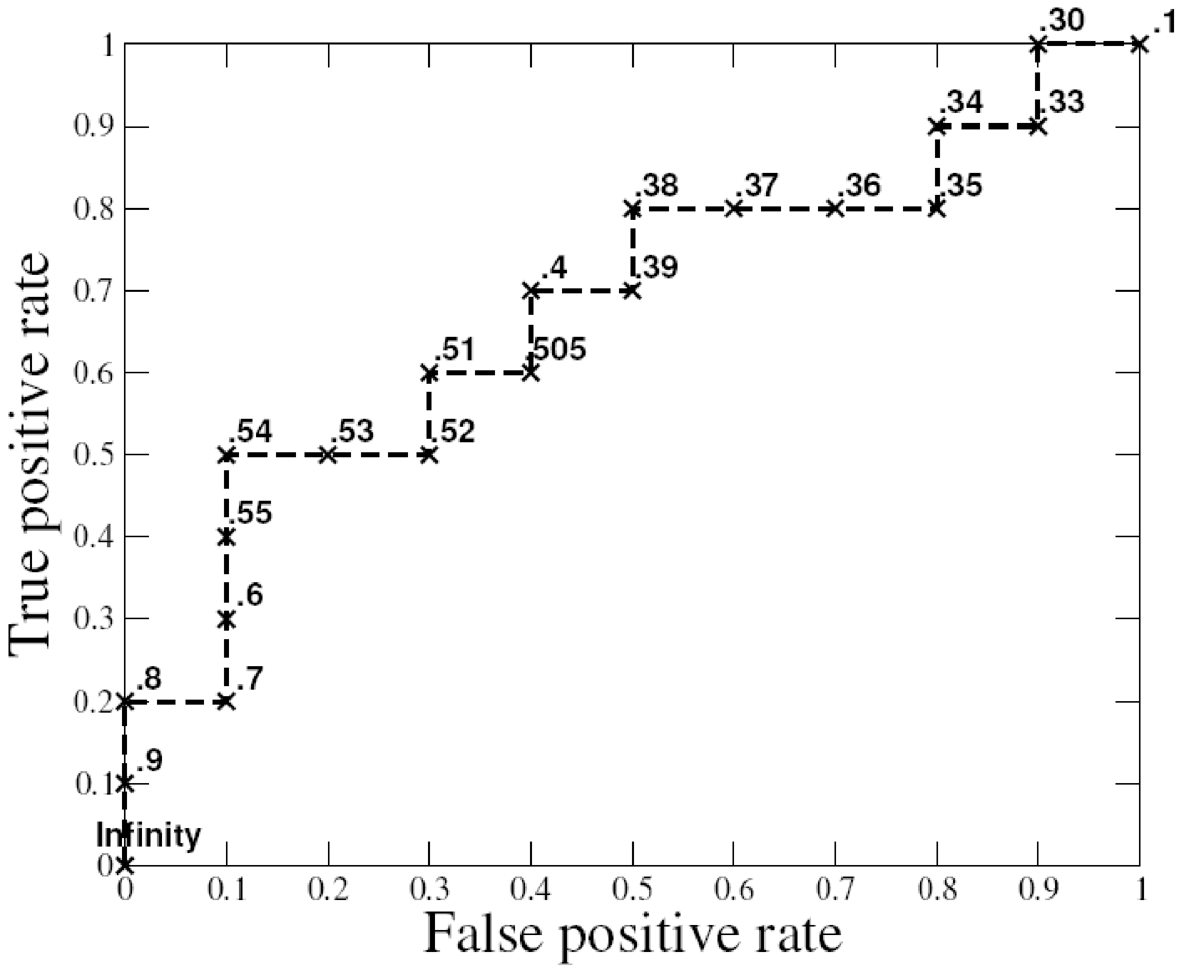
Outlook	Temp	Windy	P(Y|E)	Predicted class	Real class
overcast	mild	yes	0.95	YES	YES
rainy	mild	no	0.80	YES	YES
rainy	cool	yes	0.60	YES	NO
sunny	mild	no	0.45	YES	YES
sunny	cool	no	0.40	NO	NO
sunny	hot	no	0.35	NO	NO
sunny	hot	yes	0.25	NO	NO

The highest accuracy for this classifier is achieved at positive threshold 40%, and not with the default 50% threshold

Example

Inst#	Class	Score	Inst#	Class	Score
1	**p**	.9	11	**p**	.4
2	**p**	.8	12	**n**	.39
3	**n**	.7	13	**p**	.38
4	**p**	.6	14	**n**	.37
5	**p**	.55	15	**n**	.36
6	**p**	.54	16	**n**	.35
7	**n**	.53	17	**p**	.34
8	**n**	.52	18	**n**	.33
9	**p**	.51	19	**p**	.30
10	**n**	.505	20	**n**	.1

Example



A threshold of +inf produces the point (0; 0).

As we lower the threshold to 0.9 the first positive instance is
classified positive, yielding (0;0.1).

As the threshold is further reduced, the curve climbs up and to the right, ending up at (1;1) with a threshold of 0.1.

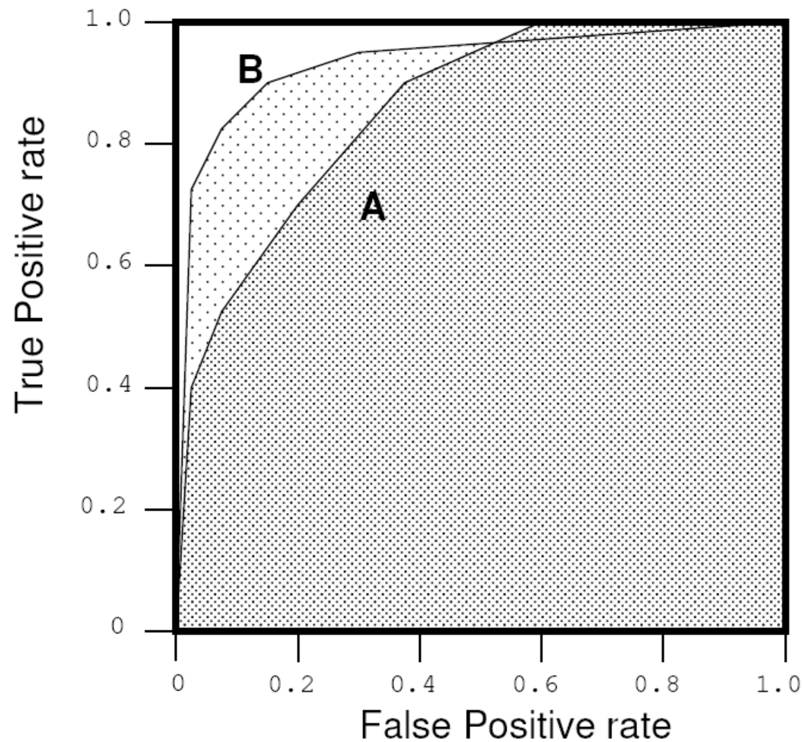
Lowering this threshold corresponds to moving from the “conservative” to the “liberal” areas of the graph.

Area under the ROC Curve (AUC)

- AUC has an important statistical property:

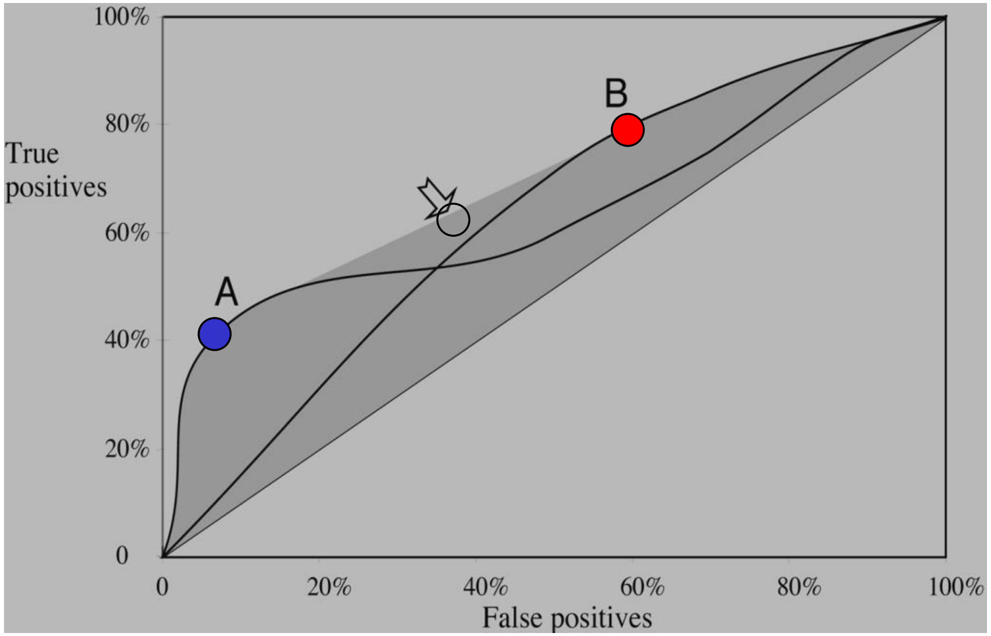
  The AUC of a classifier is equivalent to the probability that the classier will rank a randomly chosen positive instance higher than a randomly chosen negative instance.

- Often used to compare classifiers:
  - The bigger AUC the better
- AUC can be computed by a slight modification to the algorithm for constructing ROC curves.



Convex Hull

- The shaded area is called the convex hull of the two curves.
- You should operate always at a point that lies on the upper boundary of the convex hull.
- What about some point in the middle where neither A nor B lies on the convex hull?
- Answer: "Randomly" combine predictions from A and B



If you aim to cover just 40% of the true positives you should choose method A, which gives a false positive rate of 5% (expensive test, low probability of positives).

If you aim to cover 80% of the true positives you should choose method B, which gives a false positive rate of 60% as compared with A's 10%.

If you aim to cover 60% of the true positives then you should combine A and B.

Example - accuracy

2/10 correctly identified

9/10 correctly identified

Inst#	Class	Score	Inst#	Class	Score
1	**p**	.9	11	**p**	.4
2	**p**	.8	12	**n**	.39
3	**n**	.7	13	**p**	.38
4	**p**	.6	14	**n**	.37
5	**p**	.55	15	**n**	.36
6	**p**	.54	16	**n**	.35
7	**n**	.53	17	**p**	.34
8	**n**	.52	18	**n**	.33
9	**p**	.51	19	**p**	.30
10	**n**	.505	20	**n**	.1

Accuracy: 11/20

Example - accuracy

5 /10 correctly identified

9/10 correctly identified

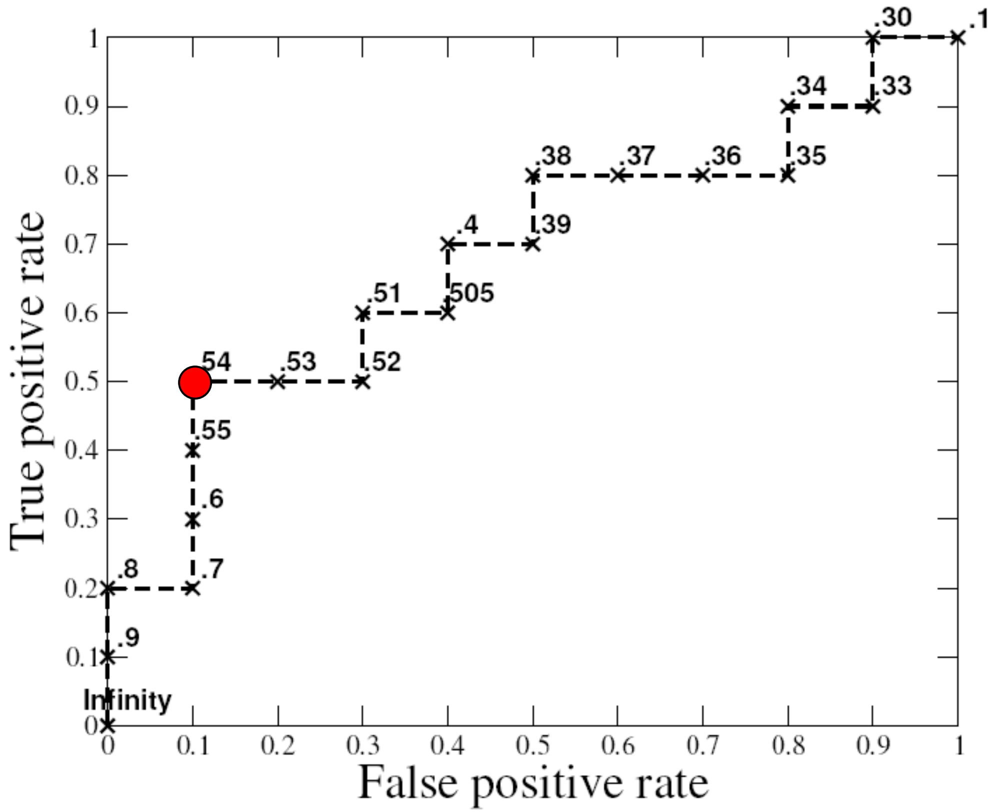
Inst#	Class	Score	Inst#	Class	Score
1	**p**	.9	11	**p**	.4
2	**p**	.8	12	**n**	.39
3	**n**	.7	13	**p**	.38
4	**p**	.6	14	**n**	.37
5	**p**	.55	15	**n**	.36
6	**p**	.54	16	**n**	.35
7	**n**	.53	17	**p**	.34
8	**n**	.52	18	**n**	.33
9	**p**	.51	19	**p**	.30
10	**n**	.505	20	**n**	.1

Accuracy: 14/20≈70%

Observations – Accuracy

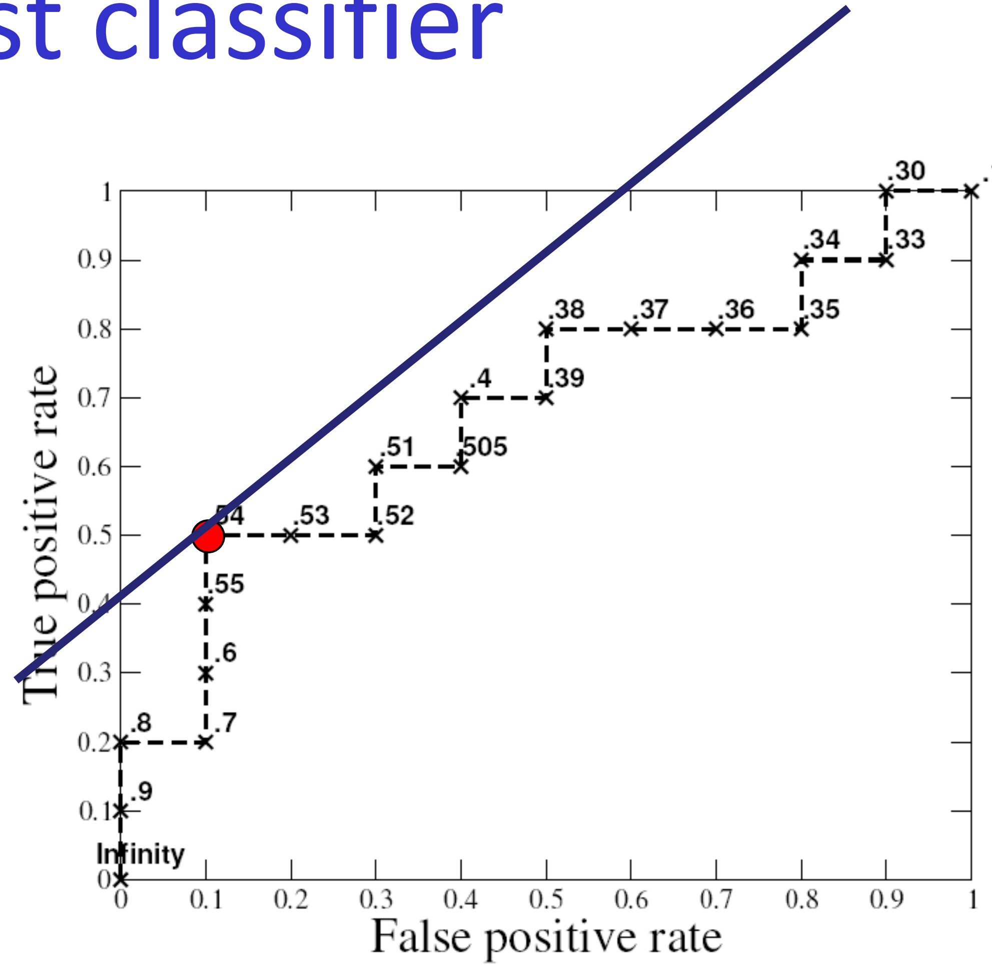
- The ROC point at (0.1, 0.5) produces its highest accuracy (14/20≈70%): identified 5 out of 10 positives and 9 out of 10 negatives correctly

- Note that the classifier's best accuracy occurs at a threshold of .54, rather than at default .5.



The best classifier

- If the costs of TP and FP are equal, then the best operating point is at the tangent of a 45^{o} line, where it touches the convex hull, and this corresponds to the maximum accuracy.

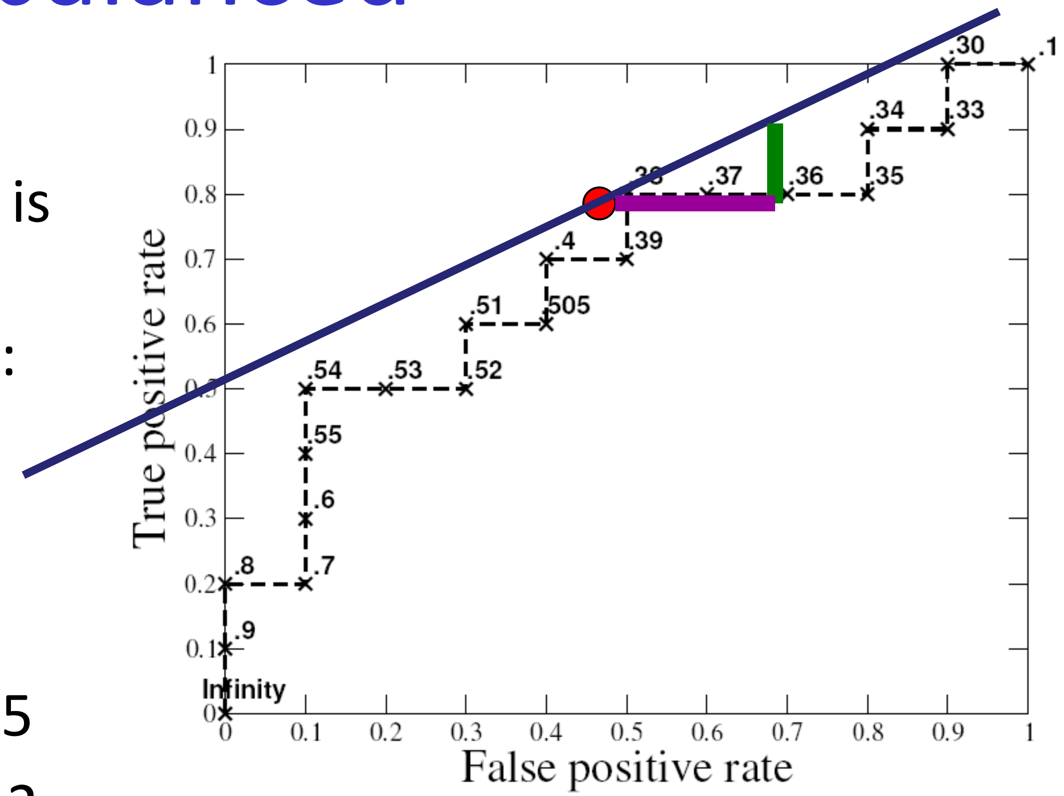


The best classifier when costs are unbalanced

- If the cost of FP and TP is different, the best classifier is at a tangent of a different line. The slope of this line is:

$$\frac{P(-)*\textcolor{green}{\textbf{cost (FP)}}}{P(+)*\textcolor{purple}{\textbf{benefit(TP)}}}$$

For example, with P(-)=P(+)=0.5 and cost (FP)=1, benefit (TP) = 2 the line will have slope 1/2



Discussion – Comparing Classifiers