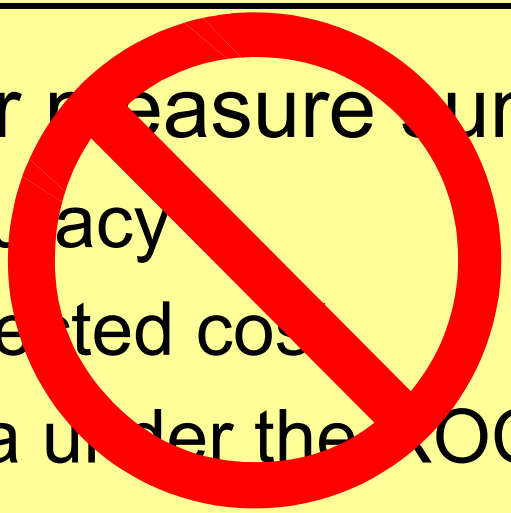
Cost Curves

Robert Holte
Computing Science Dept.
University of Alberta

Joint work with Chris Drummond, NRC, Ottawa

Cost Curve Tool programmed by Alden Flatt

How to Evaluate Performance ?



- Scalar measure summarizing performance
  - Accuracy
  - Expected cost
  - Area under the ROC curve

- Performance Visualization Techniques
  - ROC curve
  - Cost Curve

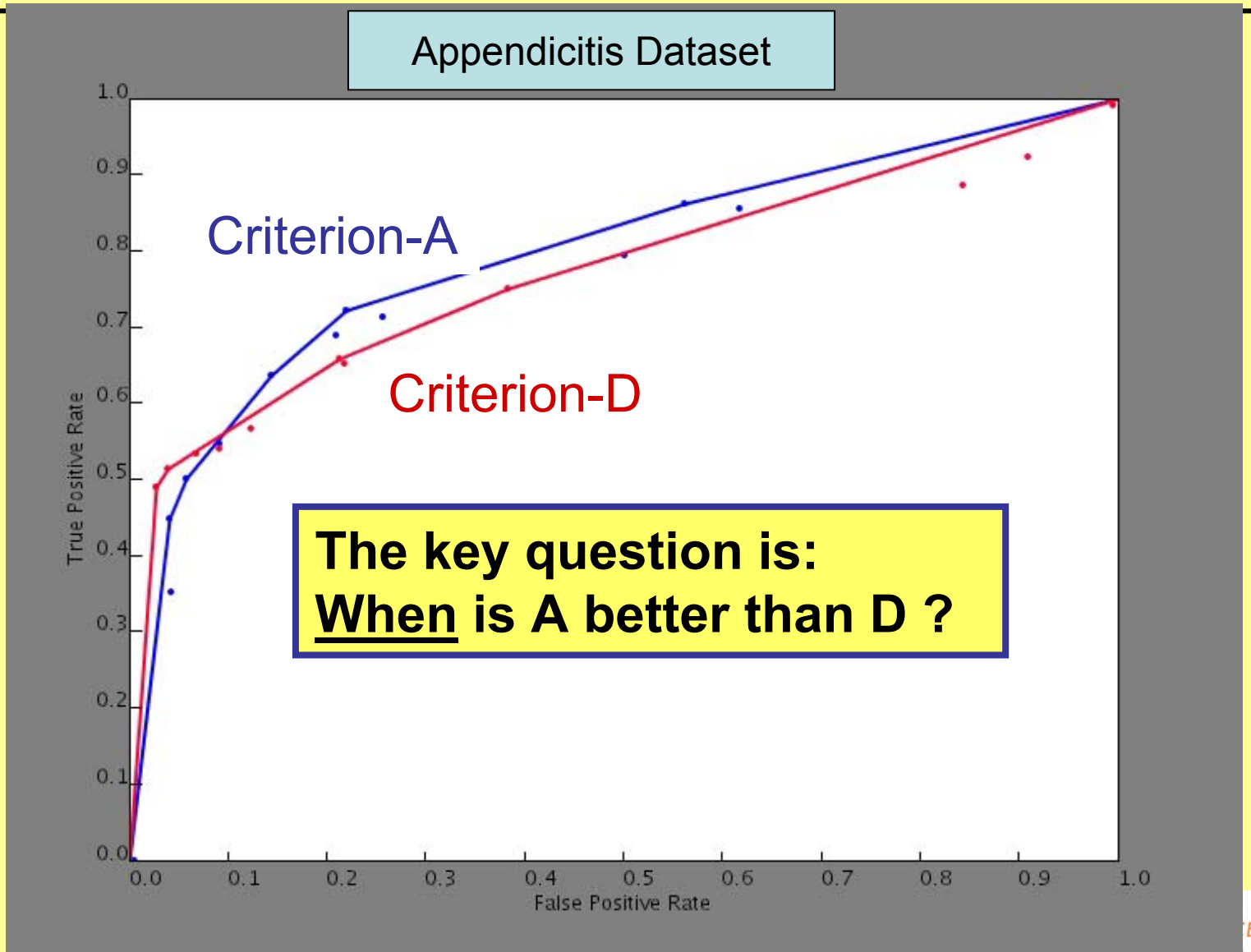
The Lure of Scalar Measures

"…it is often preferable to employ a single value measure which summarizes the performance of a classifier, e.g. because there are several classifiers to be compared and there is no clear dominance of one ROC curve above the others.

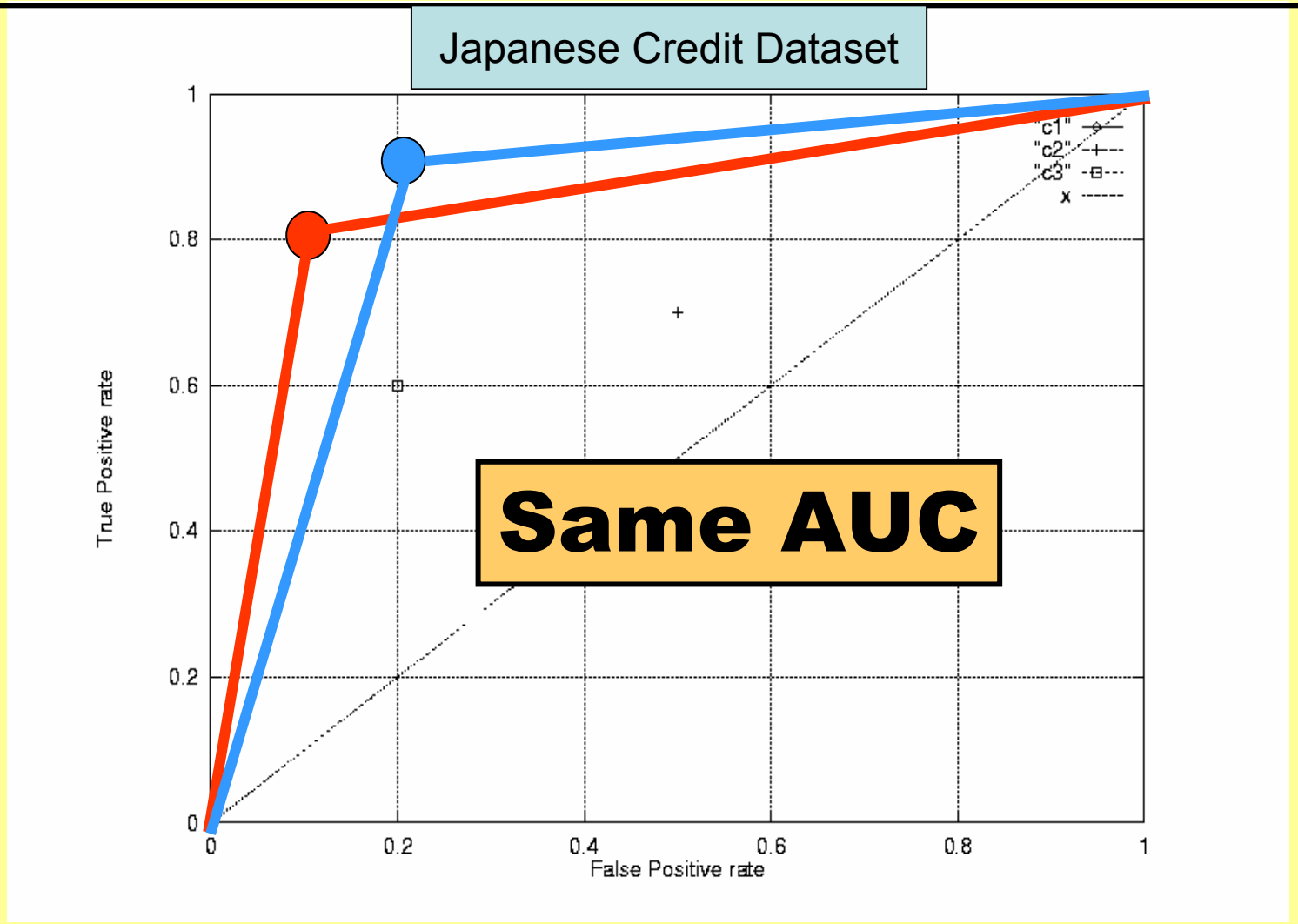
The most widely used single measure is the Area Under the ROC Curve …"

– paraphrase from a workshop paper

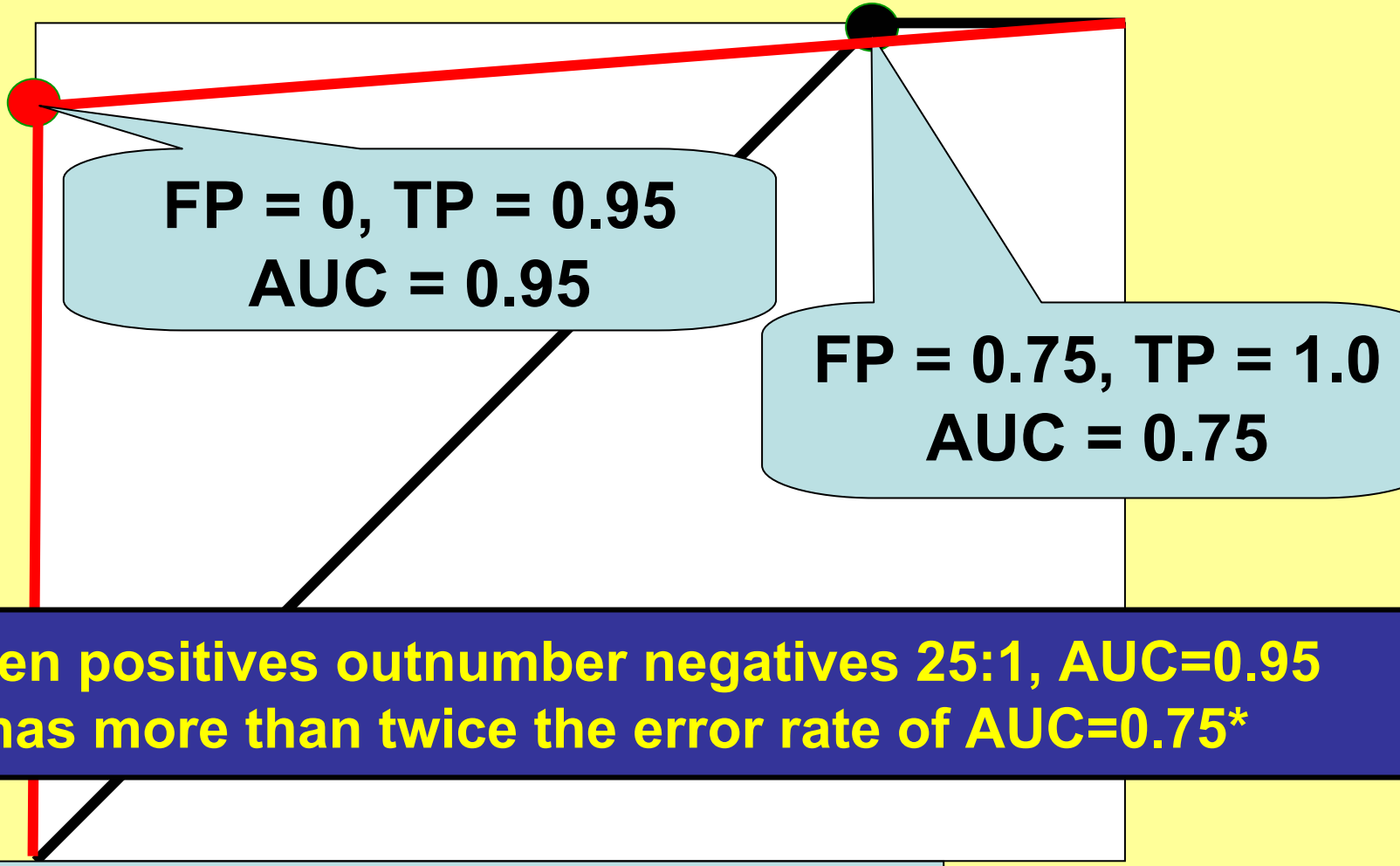
2 Splitting Criteria for C4.5



C4.5 vs 1R



Is AUC=0.95 better than AUC=0.75 ?



When positives outnumber negatives 25:1, AUC=0.95 has more than twice the error rate of AUC=0.75*

* In Phil Long's application, the ratio is 327:1

What's Genuinely Good About Scalar Measures ?

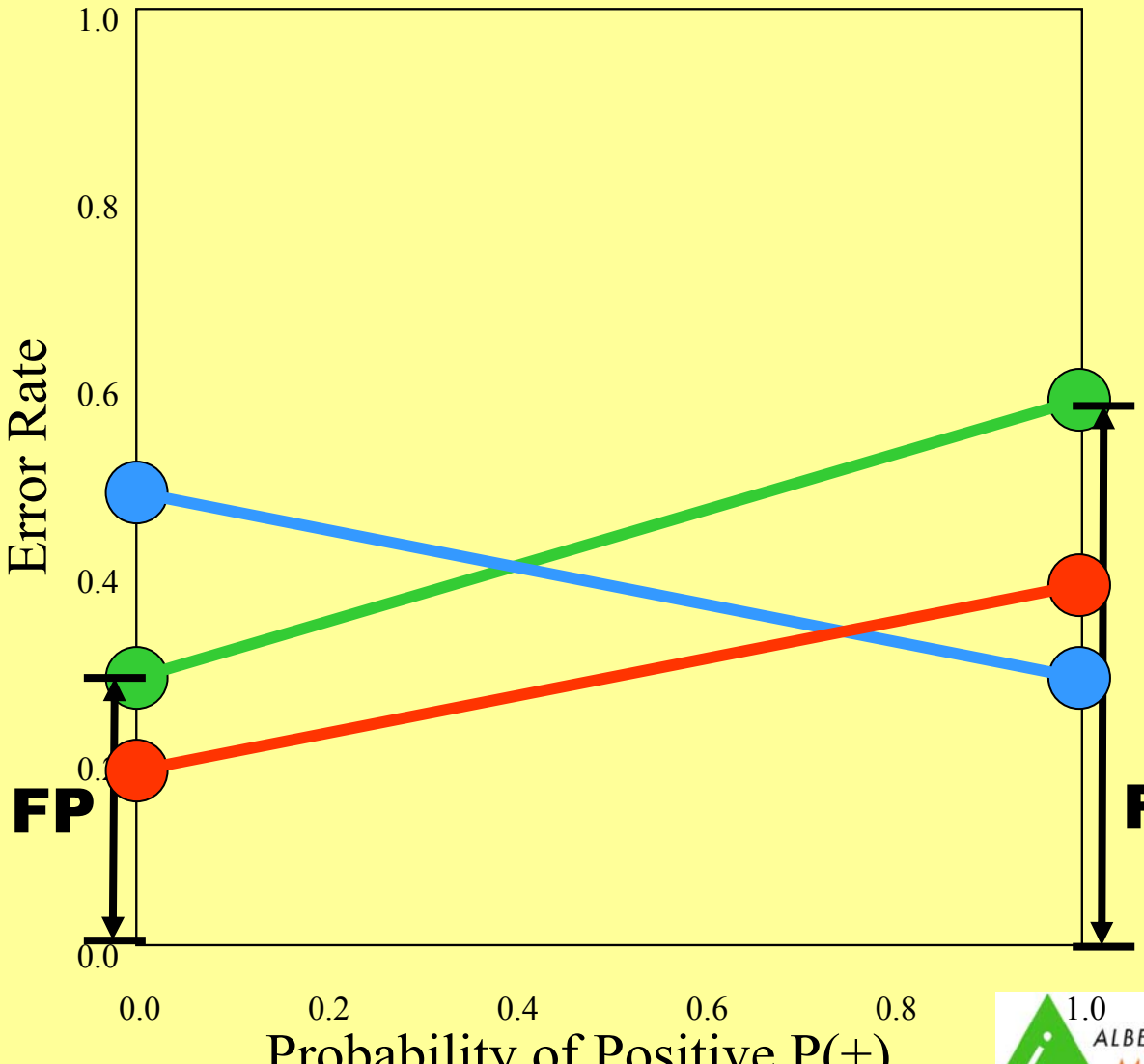
- we know how to average them, compute confidence intervals, test for significance, etc.

  … and there is off-the-shelf software to do these calculations for us.

- being one-dimensional leaves the second dimension free for other uses, e.g.
  - Learning curves
  - Multiple datasets

Cost Curves

Cost Curves

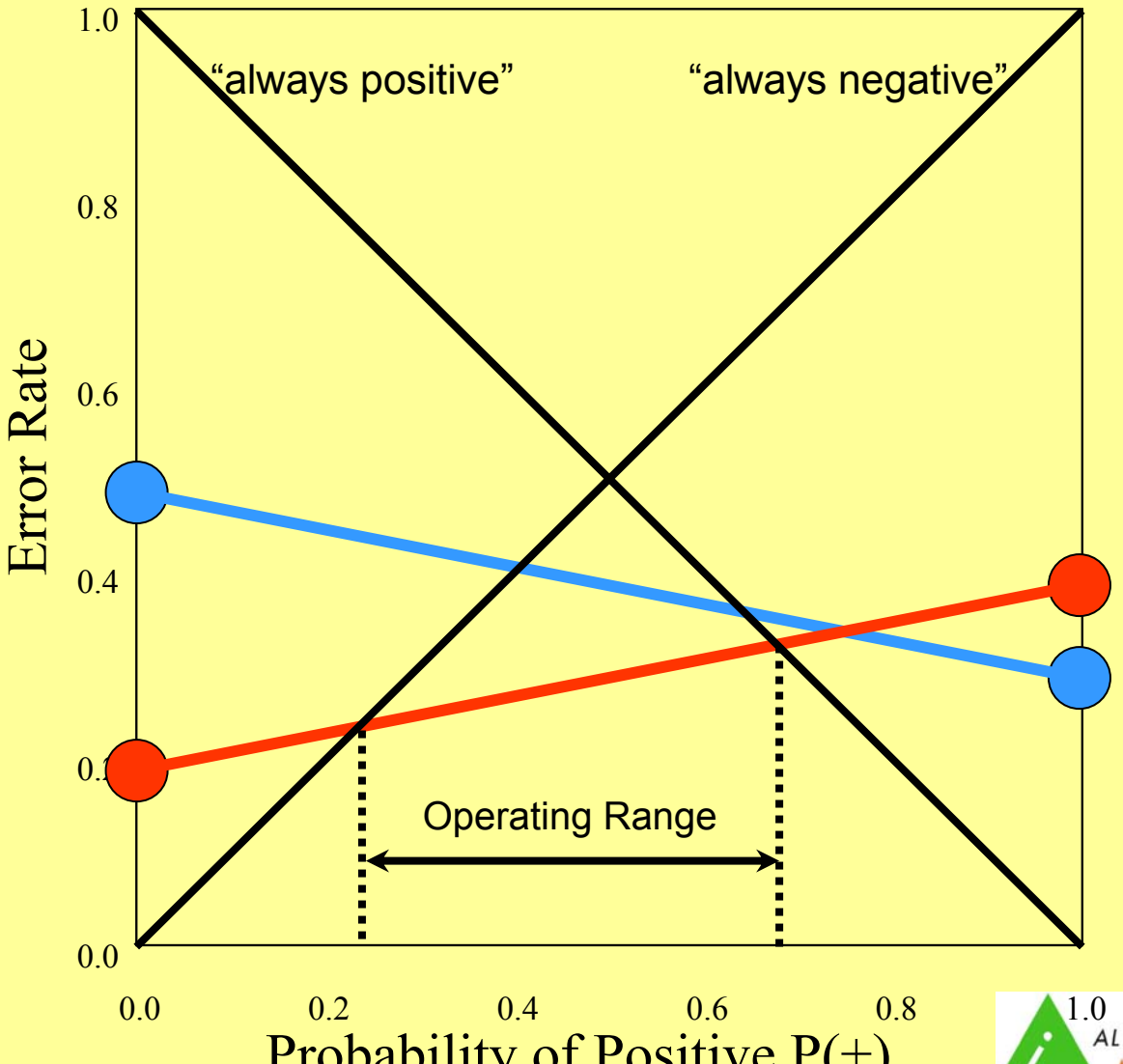


Classifier 1
TP = 0.4
FP = 0.3

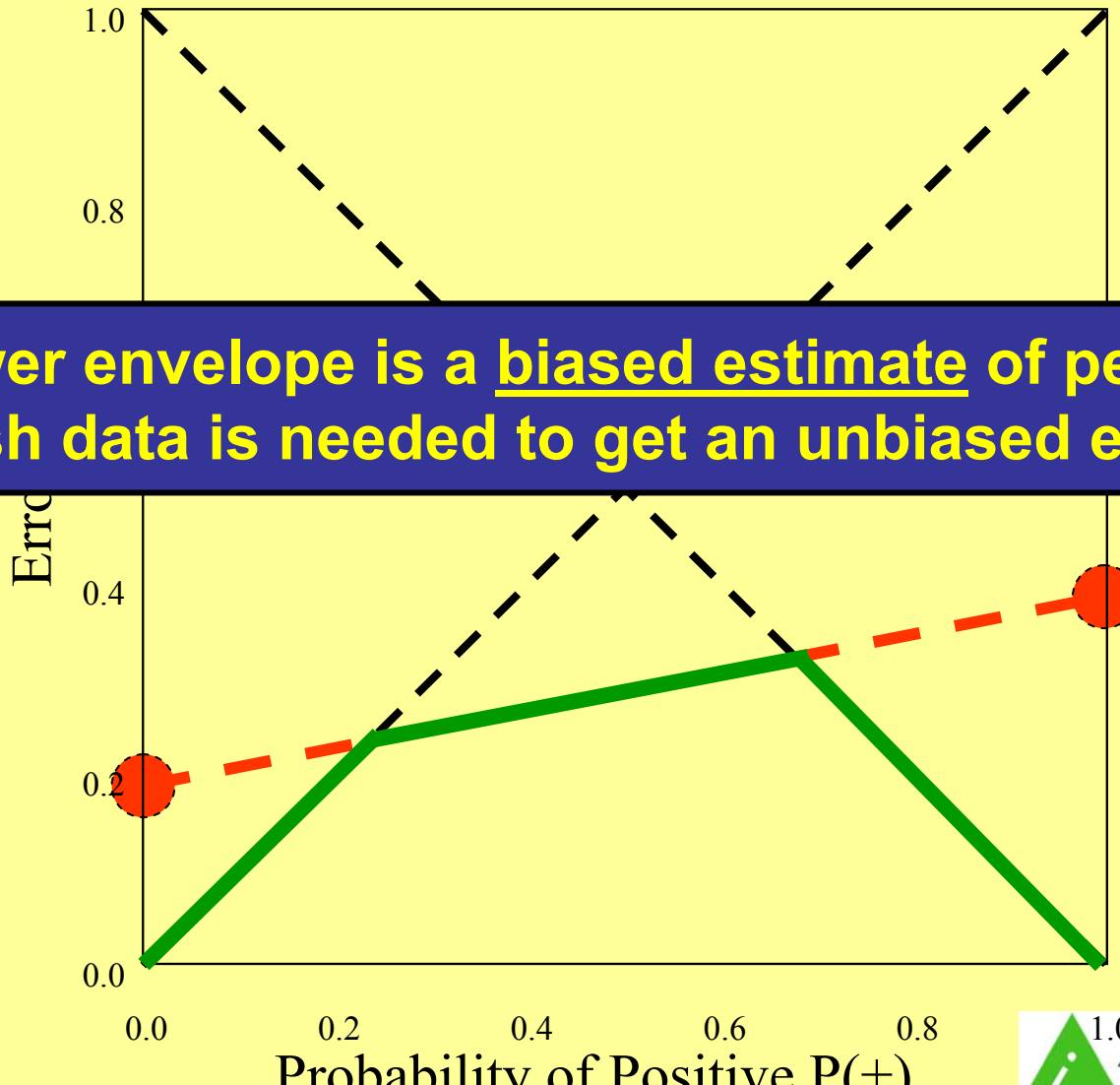
Classifier 2
TP = 0.7
FP = 0.5

Classifier 3
TP = 0.6
FP = 0.2

Operating Range

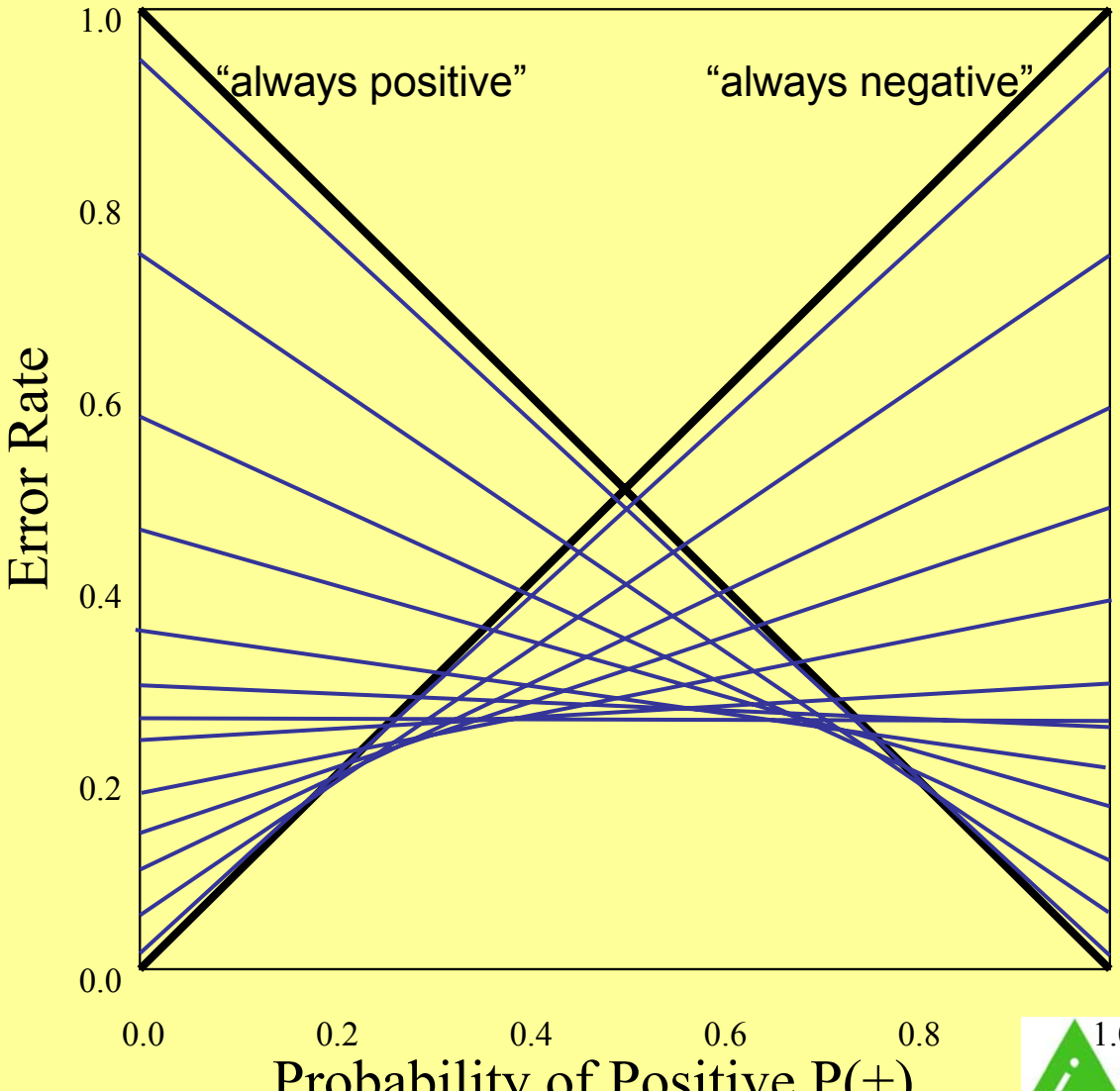


Lower Envelope



The lower envelope is a biased estimate of performance.
Fresh data is needed to get an unbiased estimate.

Varying a Threshold



Taking Costs Into Account

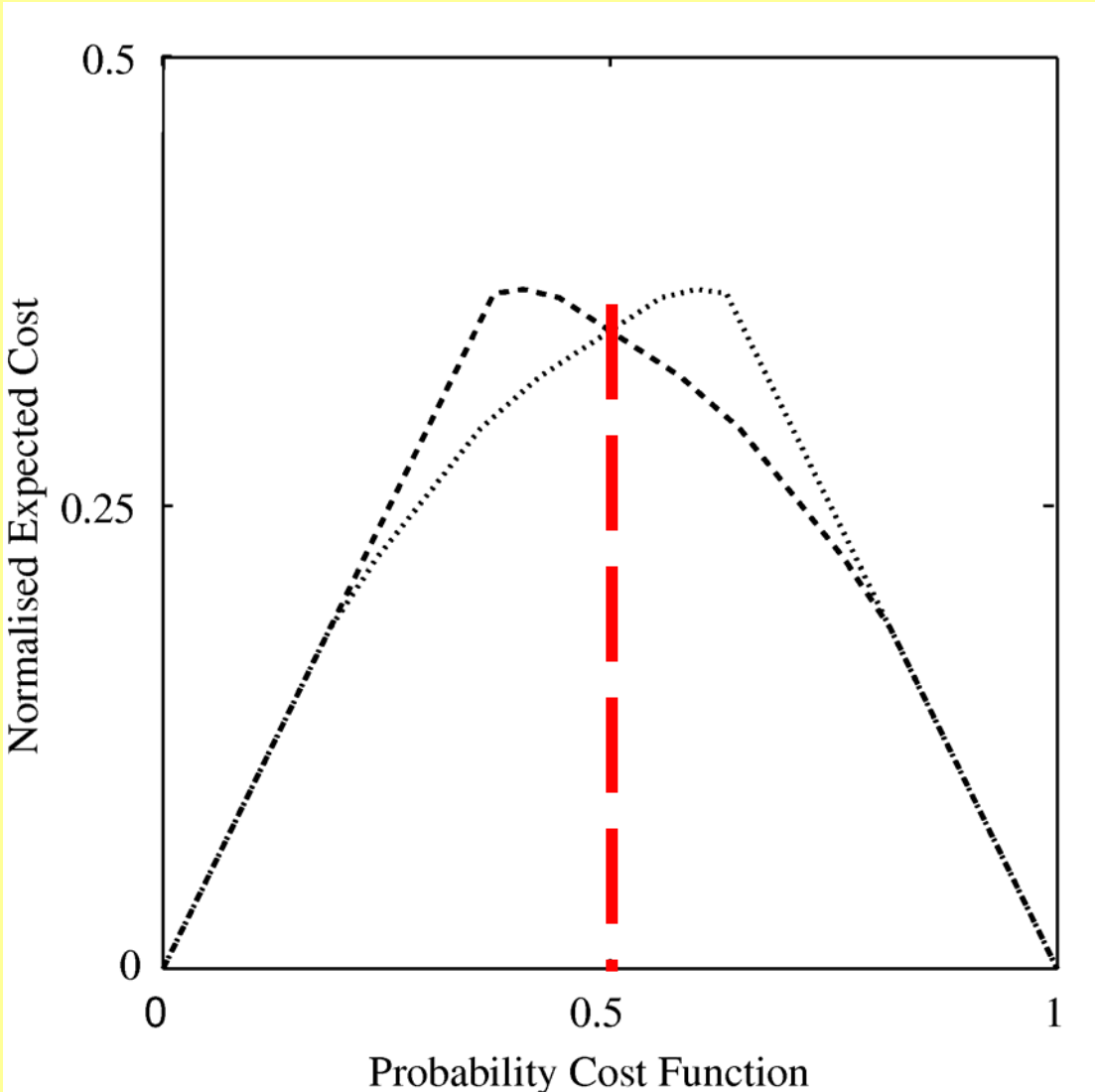
$$Y = FN \bullet X + FP \bullet (1-X)$$

So far, $X = p(+)$, making Y = error rate

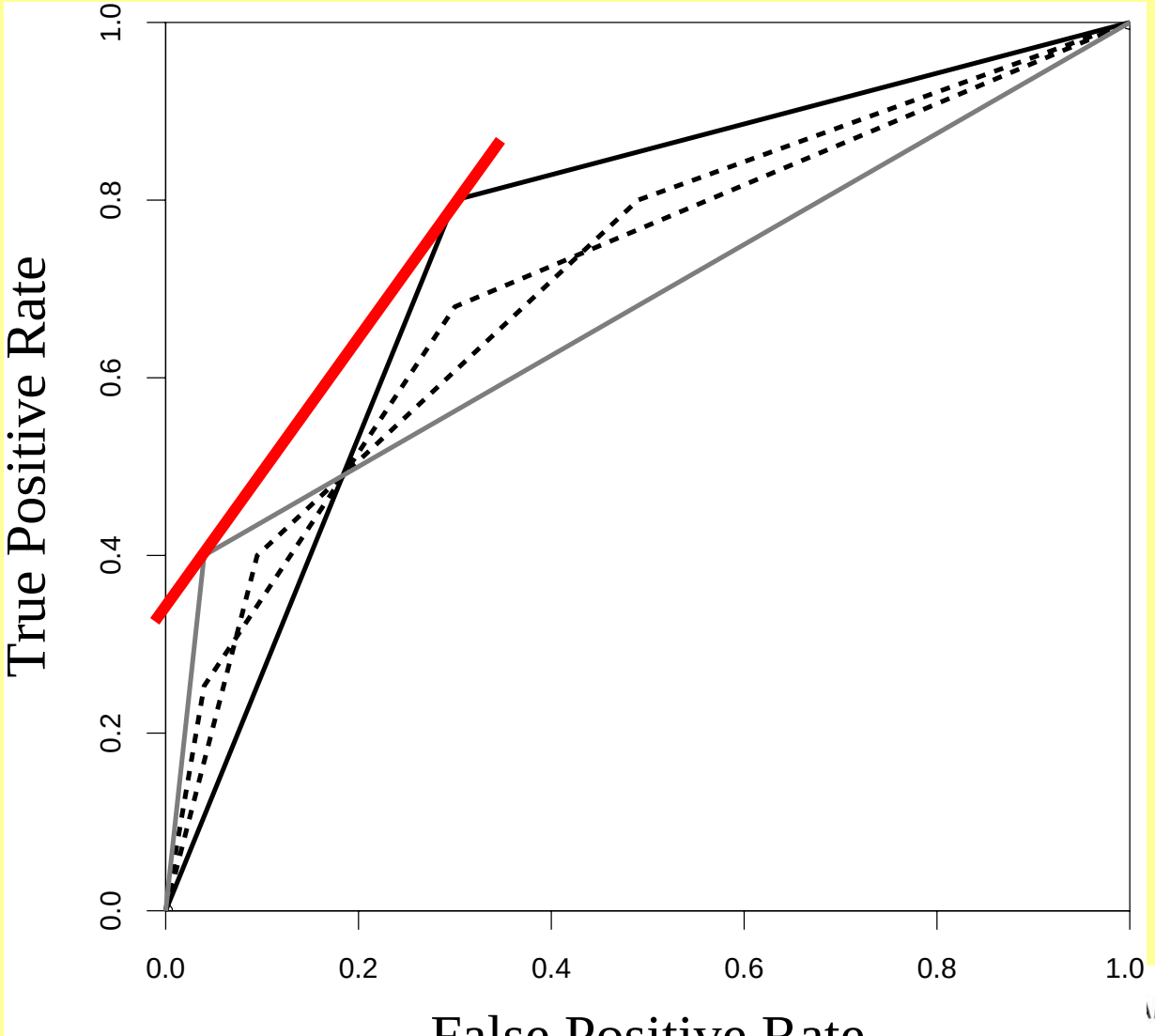
$$X = \frac{p(+) \bullet C(-|+)}{p(+) \bullet C(-|+) \; + \; (1-p(+)) \bullet C(+|-)}$$

Y = expected cost normalized to [0,1]

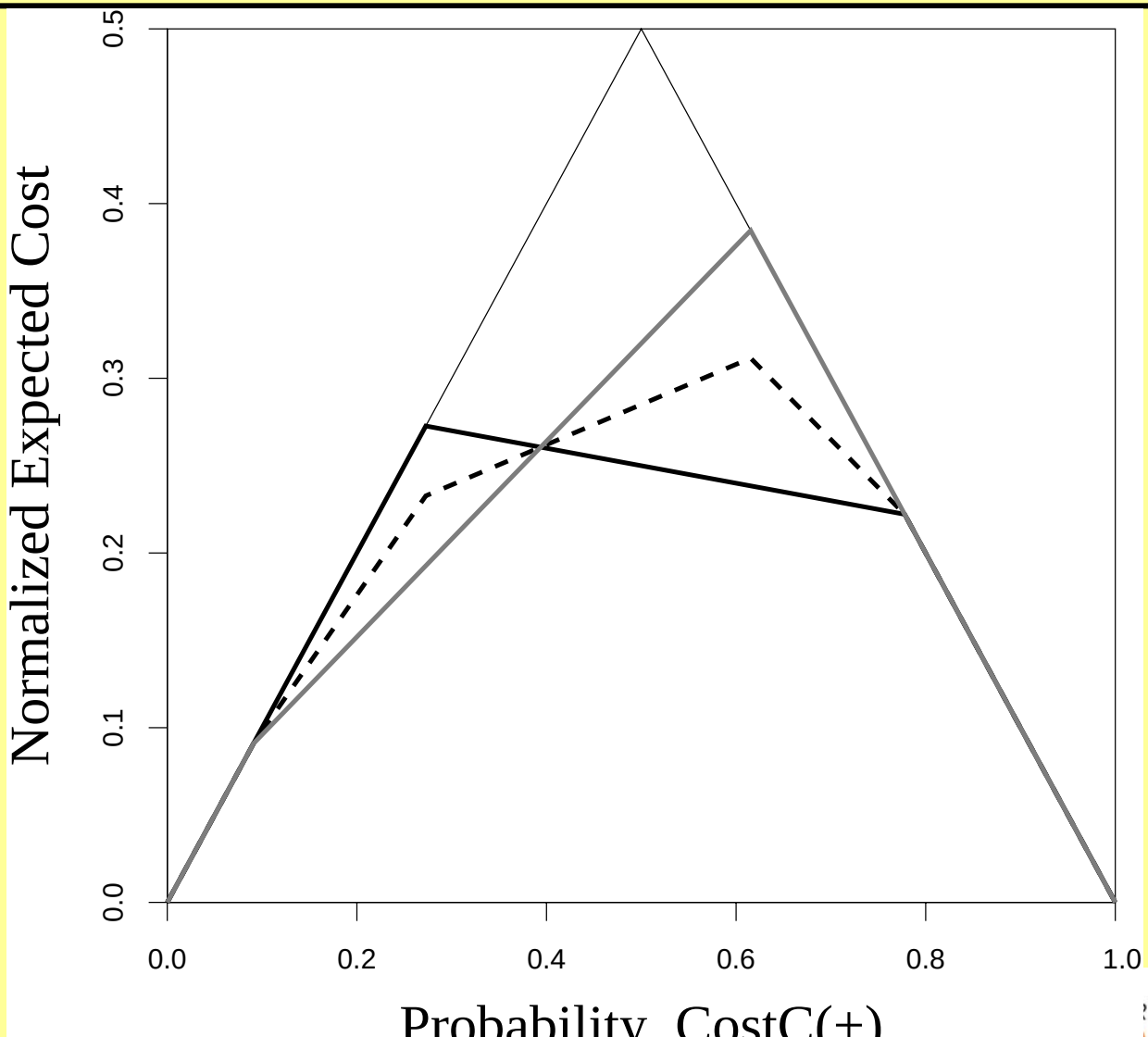
Comparing Cost Curves



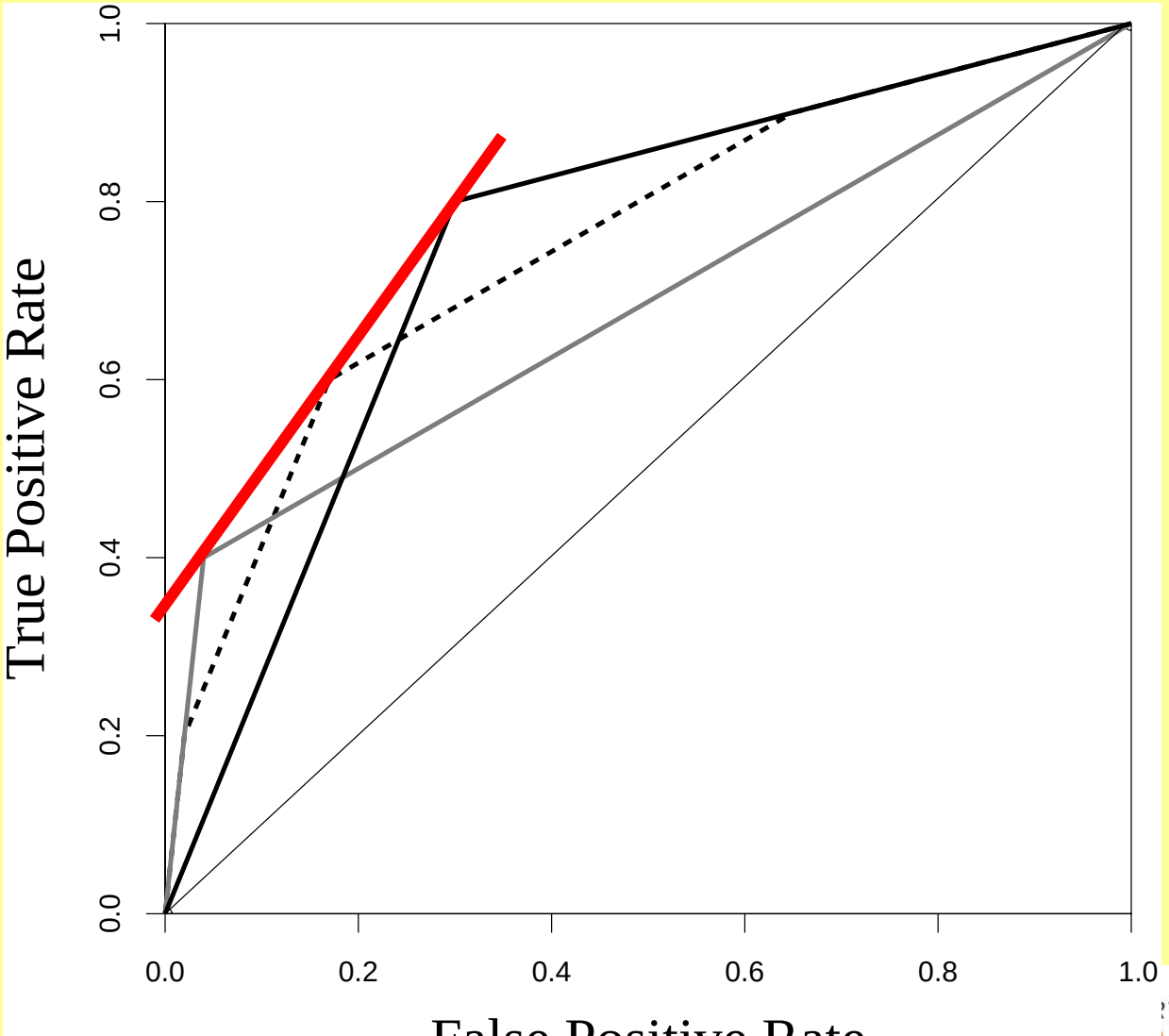
Averaging ROC Curves



Averaging Cost Curves



Cost Curve Avg. in ROC Space



Confidence Intervals

True	Predicted	
	pos	neg
pos	78	22
neg	40	60

True	Predicted	
	pos	neg
pos	75	25
neg	45	55

True	Predicted	
	pos	neg
pos	83	17
neg	38	62

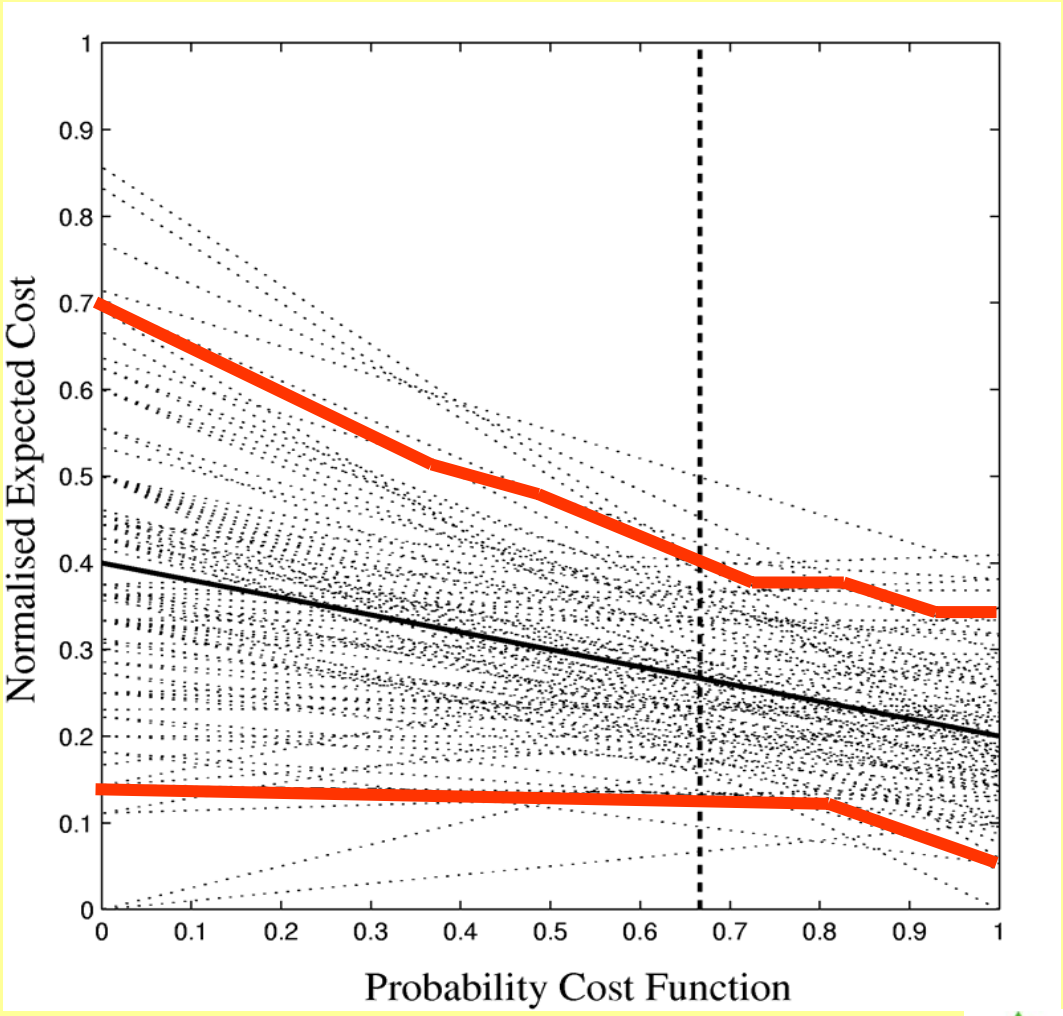
Original
TP = 0.78
FP = 0.4

Resample #1
TP = 0.75
FP = 0.45

Resample #2
TP = 0.83
FP = 0.38

Resample confusion matrix 10000 times and take 95% envelope

Confidence Interval Example



Paired Resampling to Test Statistical Significance

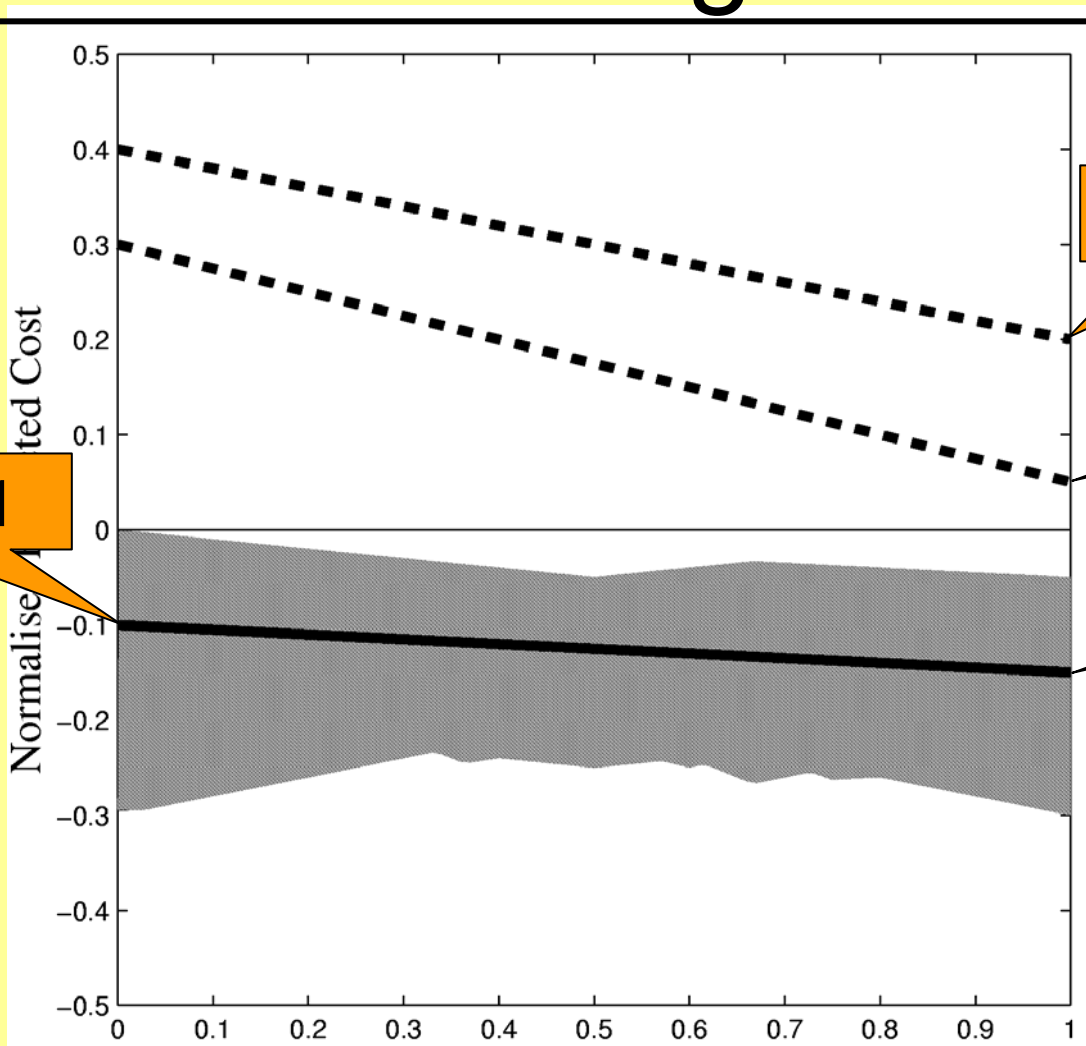
For the 100 test examples in the negative class:

Predicted by Classifier1	Predicted by Classifier2	
	pos	neg
pos	30	10
neg	0	60

FP for classifier1: (30+10)/100 = 0.40
FP for classifier2: (30+0)/100 = 0.30
FP2 – FP1 = -0.10

Resample this matrix 10000 times to get (FP2-FP1) values.
Do the same for the matrix based on positive test examples.
Plot and take 95% envelope as before.

Paired Resampling to Test Statistical Significance



classifier1

classifier2

FP2-FP1

FN2-FN1

Correlation between Classifiers

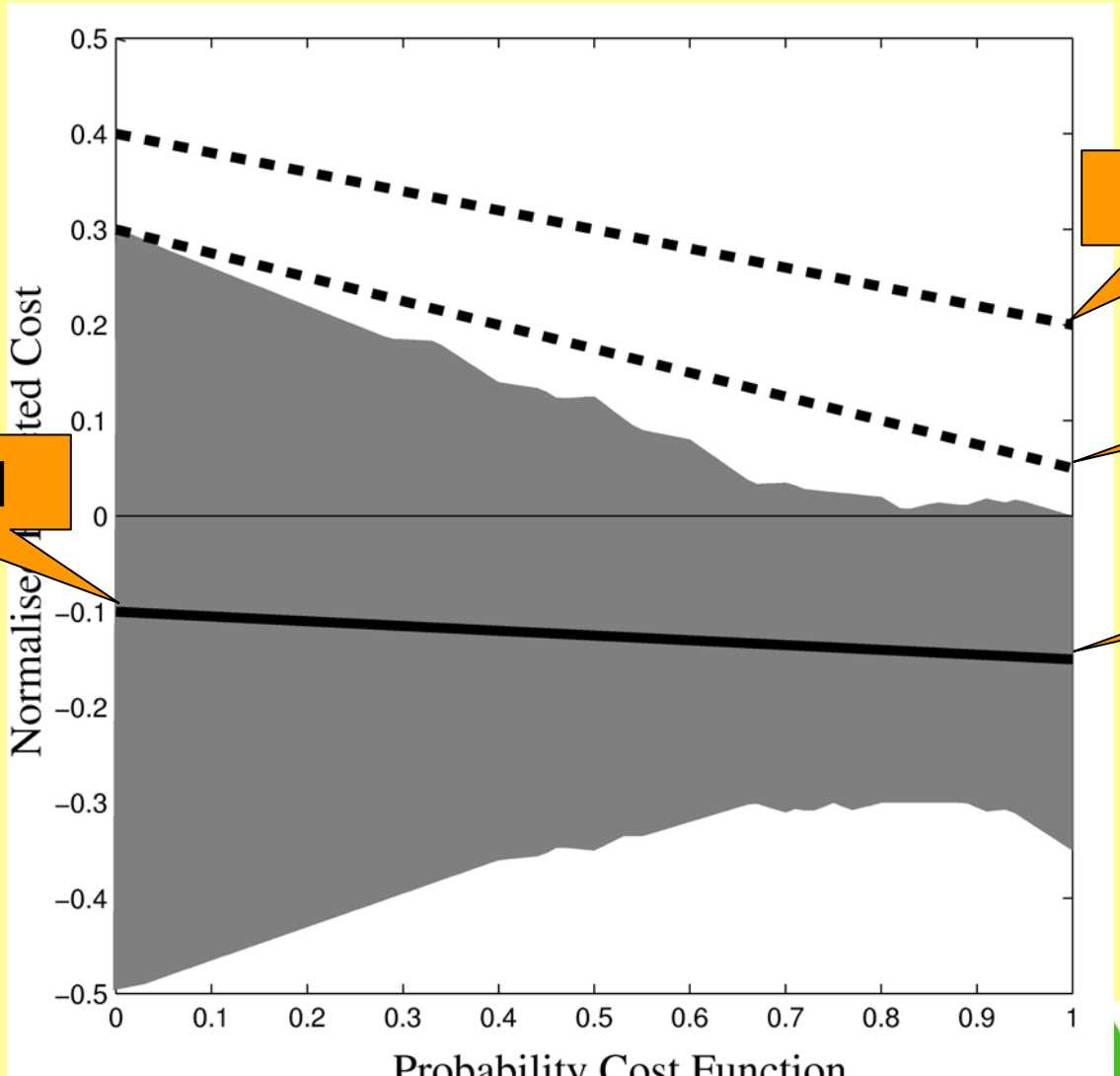
High Correlation

Predicted by Classifier1	Predicted by Classifier2	
	pos	neg
pos	30	10
neg	0	60

Low Correlation (same FP1 and FP2 as above)

Predicted by Classifier1	Predicted by Classifier2	
	pos	neg
pos	0	40
neg	30	30

Low correlation = Low significance



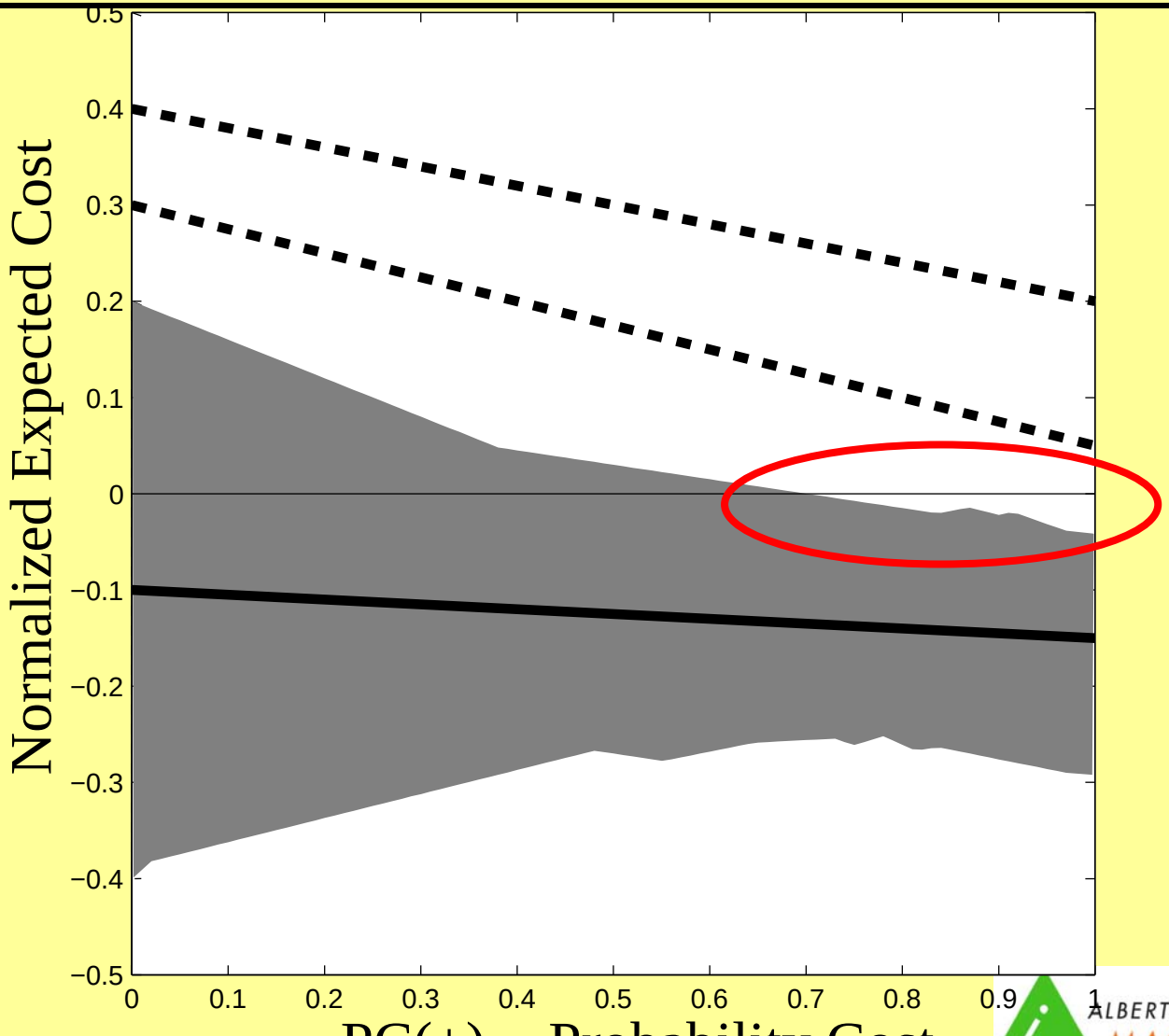
classifier1

classifier2

FP2-FP1

FN2-FN1

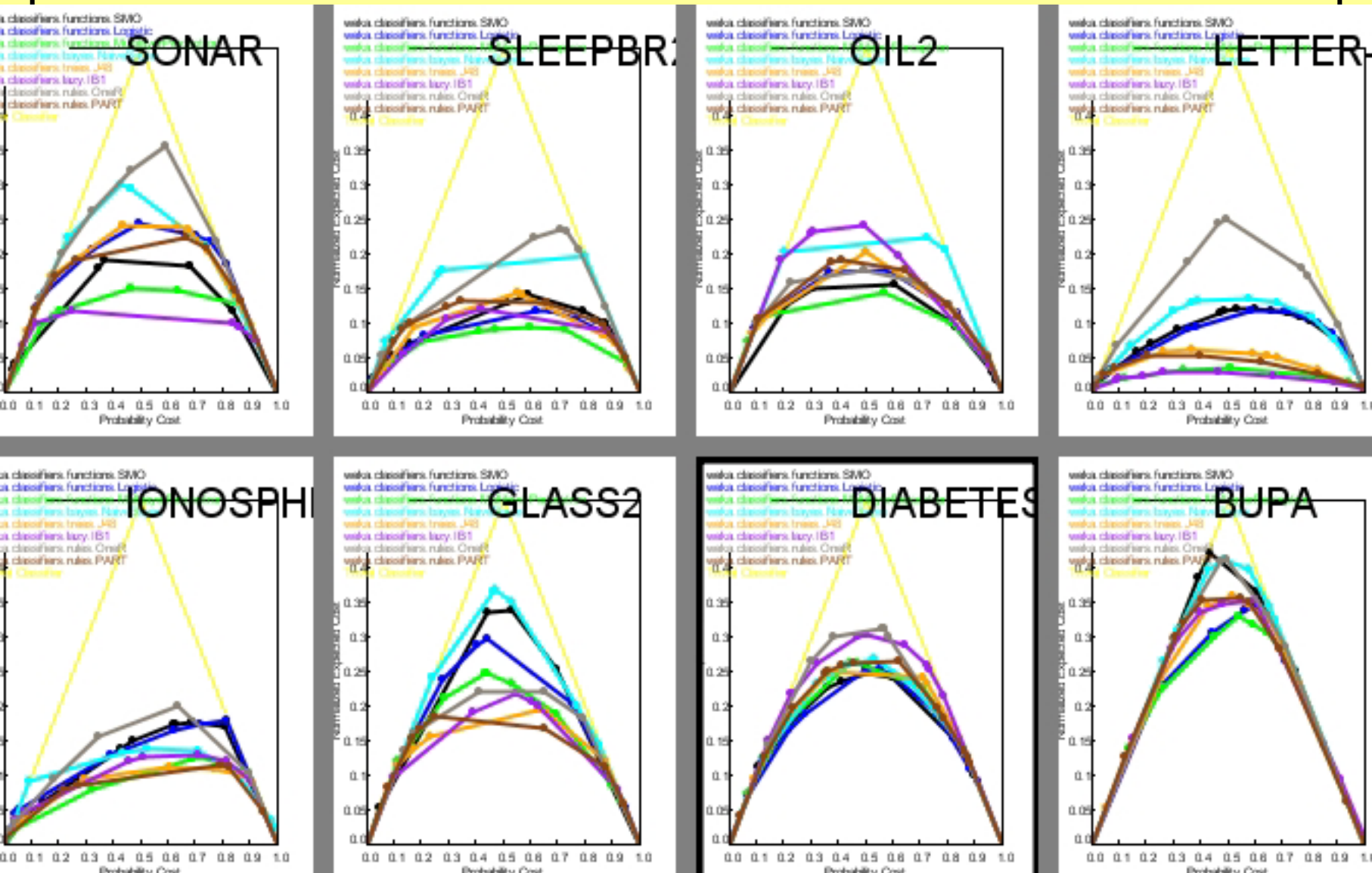
Limited Range of Significance



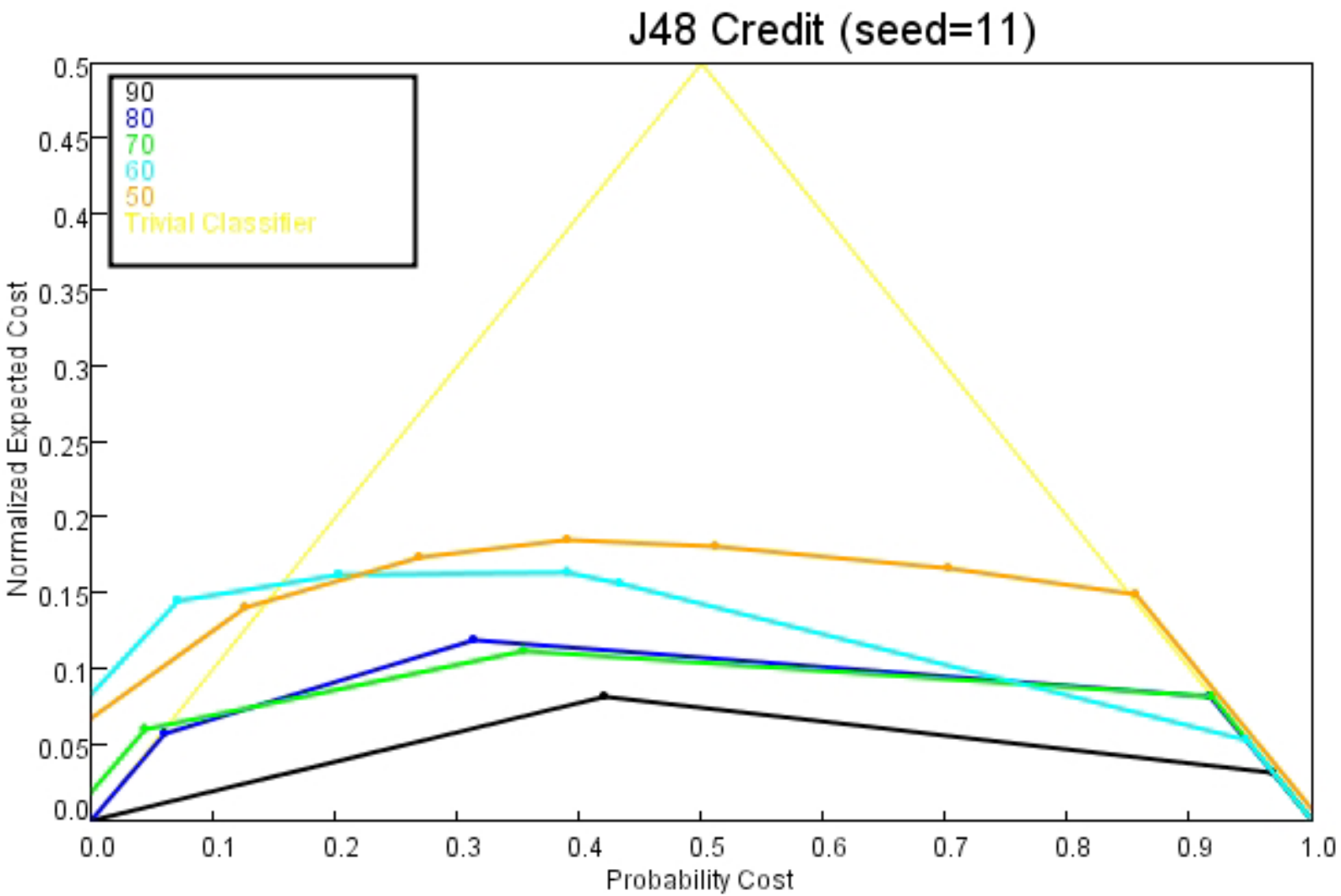
Comparing J48 and AdaBoost

Lower Envelope is Biased

Multiple Alg/Dataset Comparisons

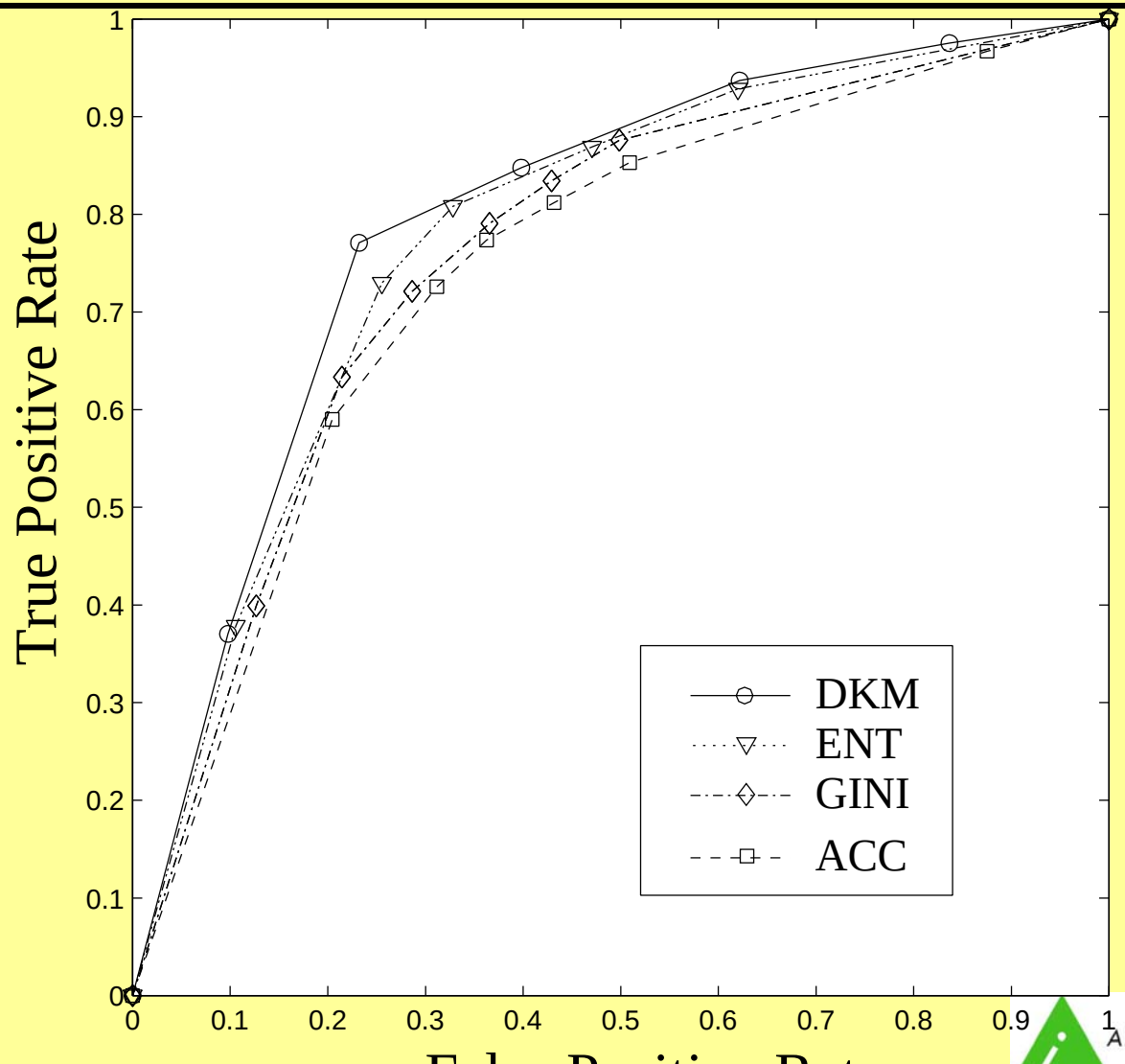


Learning Curves

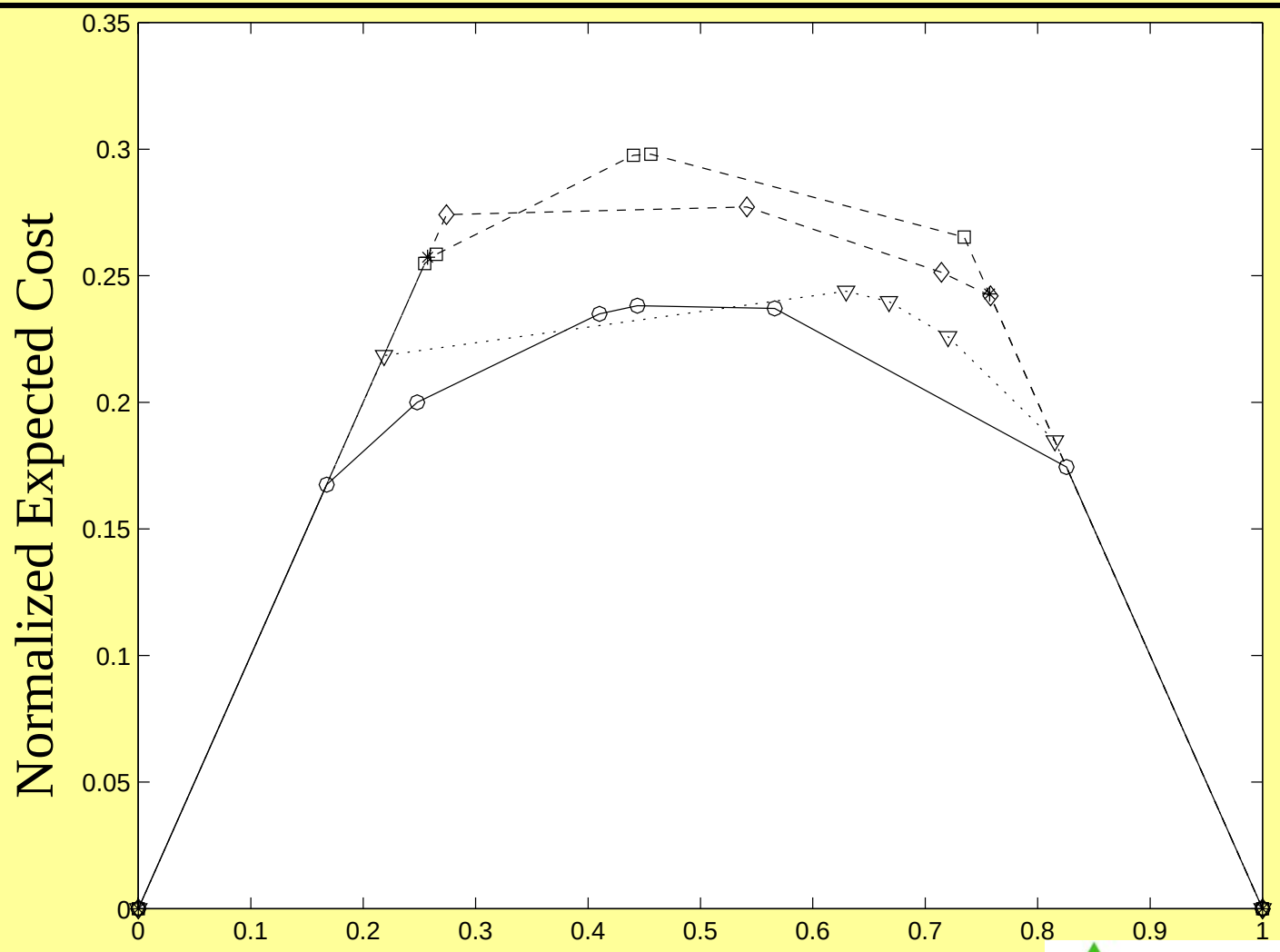


Better Data Analysis

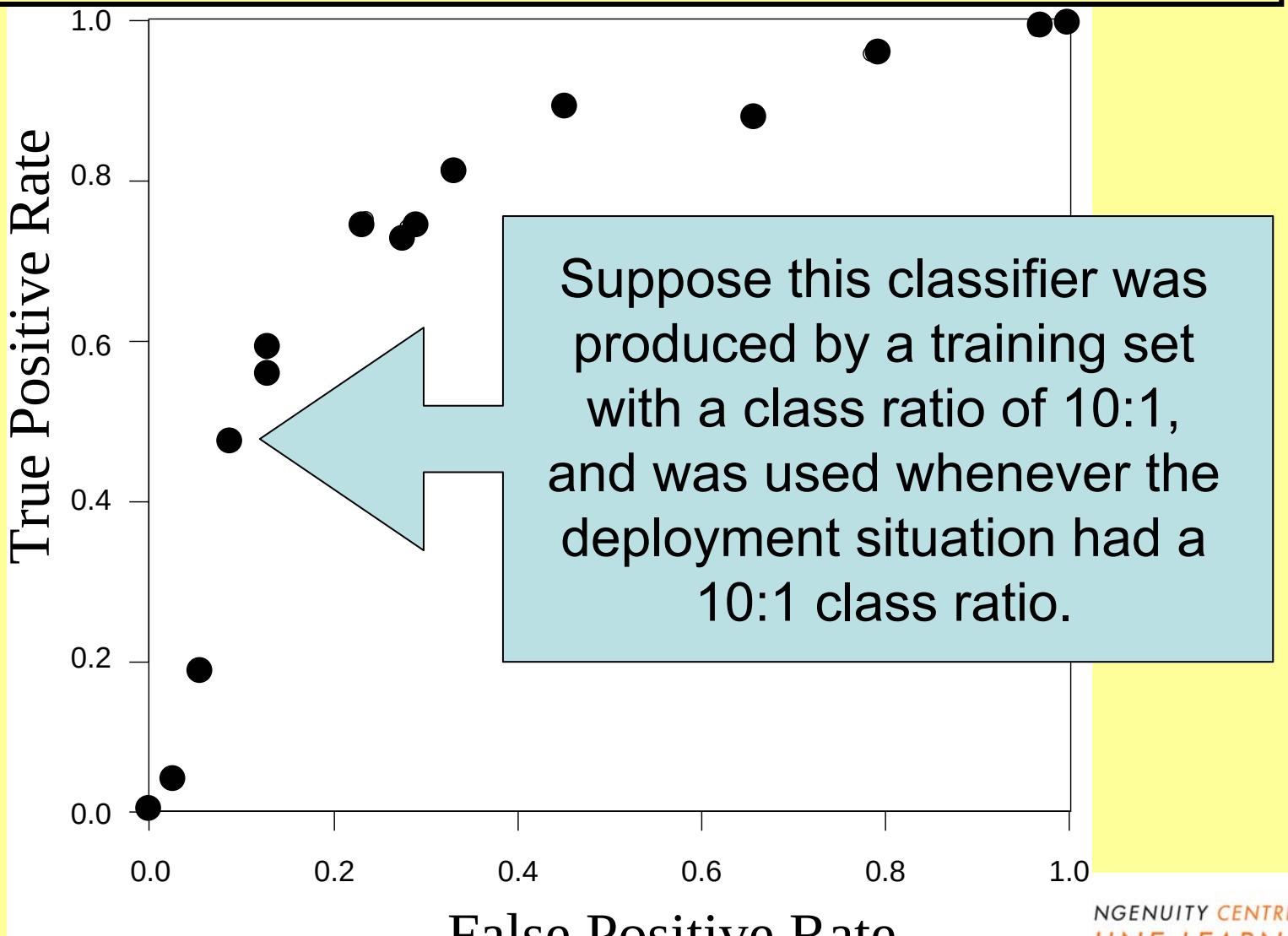
ROC, C4.5 Splitting Criteria



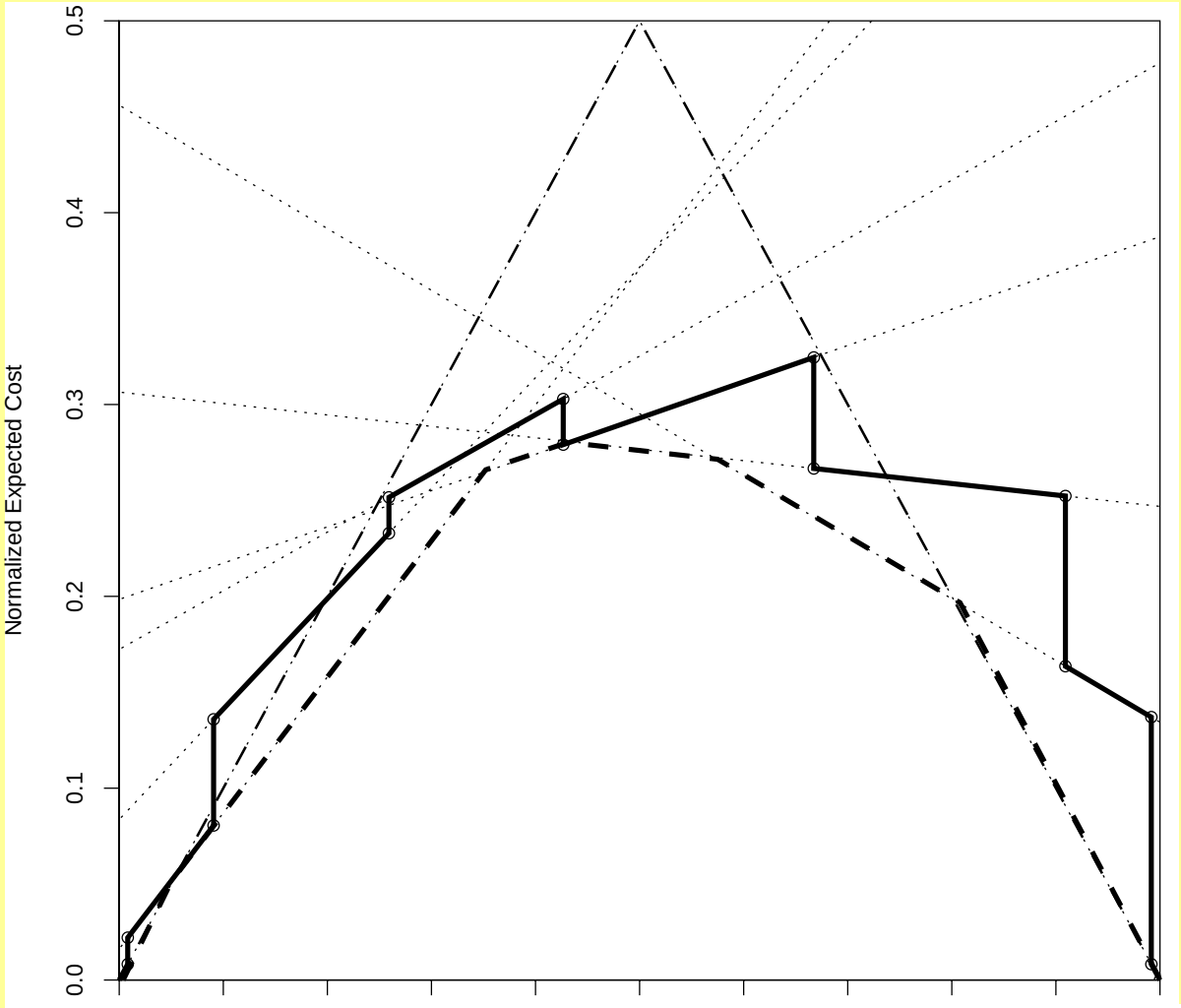
Cost Curve, C4.5 Splitting Criteria



ROC, Selection procedure



Cost Curves, Selection Procedure



Conclusions

- Scalar performance measures, including AUC, do not indicate when one classifier is better than another.
- Cost curves enable easy visualization of
  - Average performance (expected cost)
  - operating range
  - confidence intervals on performance
  - difference in performance and its significance
- Cost/ROC curve software is available. Contact: holte@cs.ualberta.ca