Model Evaluation

Myra O'Regan

October 18, 2017

Continuing with the Classification tables

Cutoff=0.3

		Predicted		
		+	-	Total
Actual	+	3	1	4
	-	4	2	6

Sensitivity(TPR)=0.75 or 75%;
Specificity(TNR)= 0.33 or 33%;
False positive (FPR)=0 .67 or 67%

Classification tables

Cutoff=0.65

		Predicted		
		+	-	Total
Actual	+	2	2	4
	-	2	4	6

Sensitivity(TPR)=0.5 or 50%;
Specificity(TNR)= 0.6 or 67%;
False positive (FPR)= 0.33 or 33%

Classification tables

Cutoff=1.00

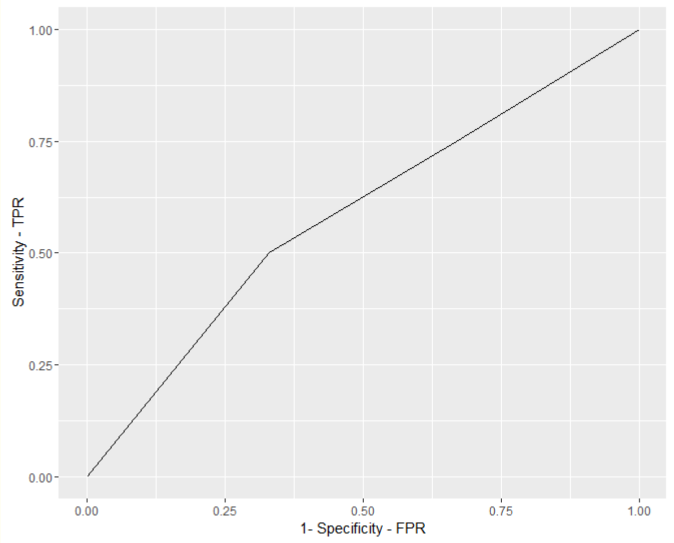
		Predicted		
		+	-	Total
Actual	+	0	4	4
	-	0	6	6

Sensitivity(TPR)=0;
Specificity(TNR)= 1 or 100%;
False positive (FPR)= 0

Results

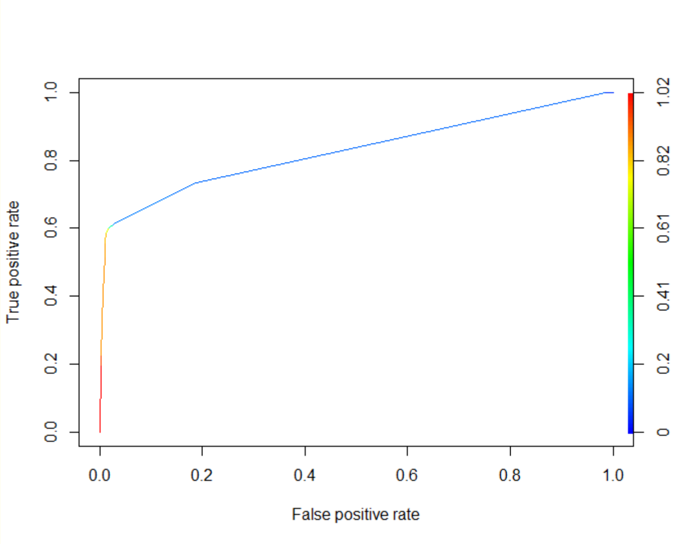
Splits	TPR	TNR	FPR
1	0	1	0
0.65	0.5	0.67	0.33
0.3	0.75	0.33	0.67
0	1	0	1

ROC curve



1.00
0.75
0.50
0.25
0.00
Sensitivity - TPR
0.00
0.25
0.50
0.75
1.00
1- Specificity - FPR

ROC for churn data



ROC curve

- 2 dimensional plot TPR on y-axis and FPR on x-axis
- A single confusion matrix produces a single point
- ROC curve formed by connecting a series of such points including (0,0) and (1,1)
- Depicts tradeoffs between benefits (true positives) and costs (false positives)
- Trivial classifiers
- Only an estimate

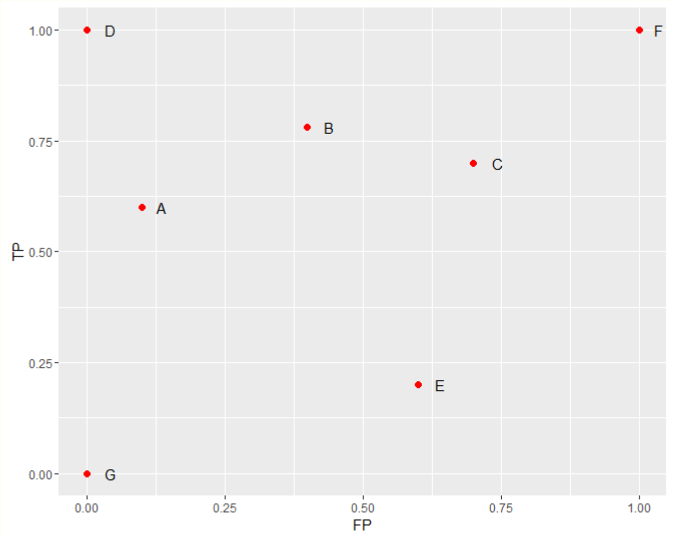
Some more on the ROC

- What is the optimal point?
- Trivial Classifiers - where are they
- Area under Curve (AUC)
- Measures overall performance
- How about uncertainty?

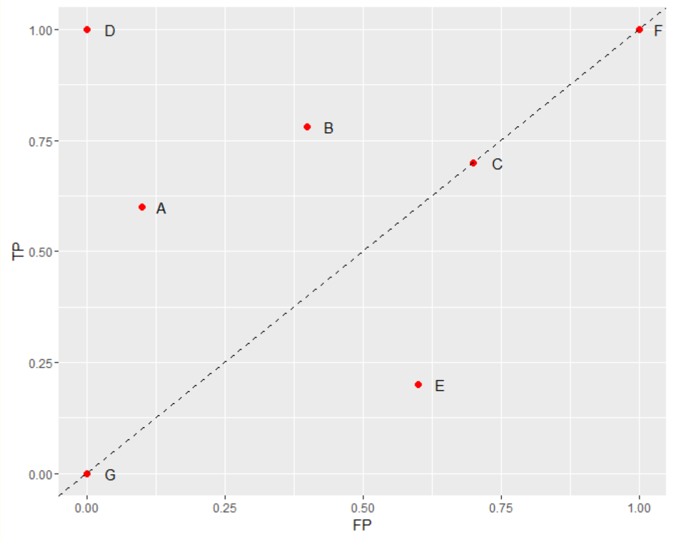
Another interpretation of the AUC

- AUC is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance
- Closely related to the Mann Whitney test
- http://www.dataschool.io/roc-curves-and-auc-explained/

7 classifiers



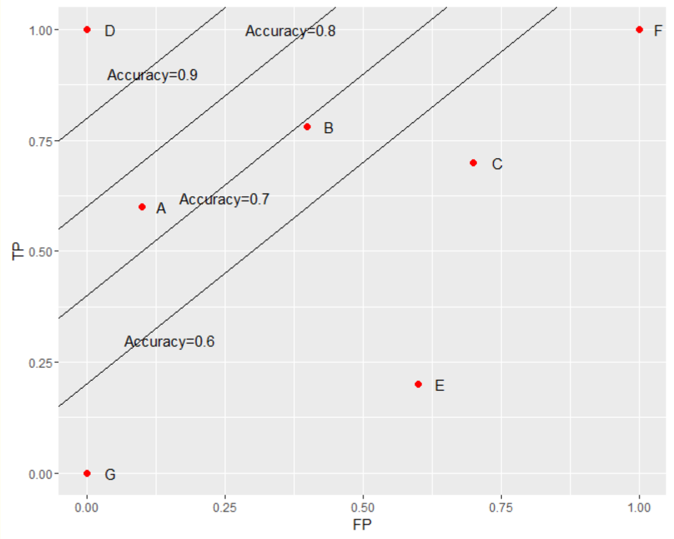
7 classifiers with random line



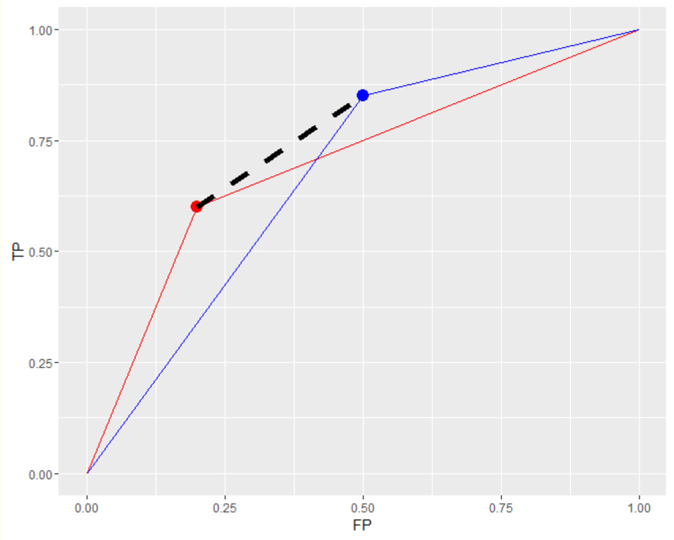
Lines in ROC Space

- TPR and TNR are not affected by prior probabilities
- p(+) : prior probabilites of +; p(-) of - =1-p(+)
- What about Accuracy?
- Acc=p(+) *TPR+p(-)*(1-FPR)
- Can write this as follows $TPR = \frac{Acc - p(-)}{p(+)} + \frac{p(-)}{p(+)} * FPR$

Isoperformance lines



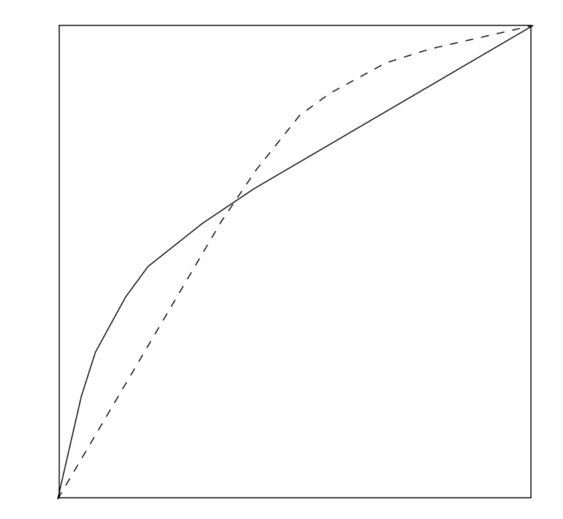
2 classifiers



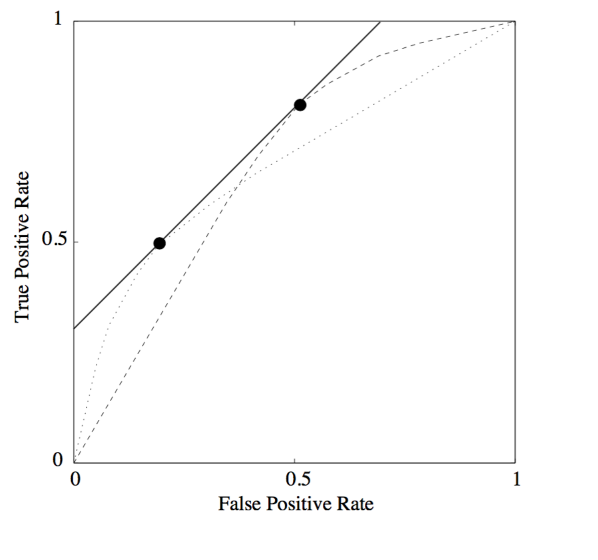
What does it mean

- Blue classifier - 2 slopes
- Define the range of ratio of $\frac{p(-)}{p(+)}$ for which classifier is useful.
- The same for red line
- slope(dashed line) class distributions for which red classifier is the same as the blue classifier.

2 ROC curves



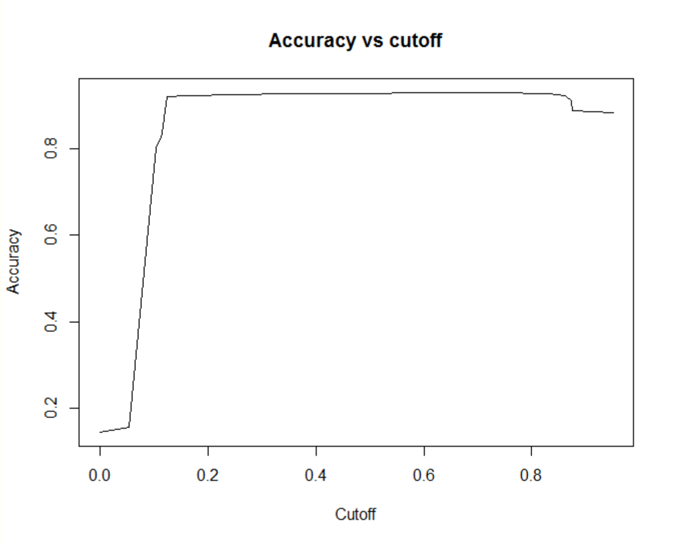
2 classifiers



What does this mean

- The solid line is the iso-performance line tangent to both two ROC curves.
- Its slope represents the operating point $\frac{p(-)}{p(+)}$ at which the two classifiers have equal performance

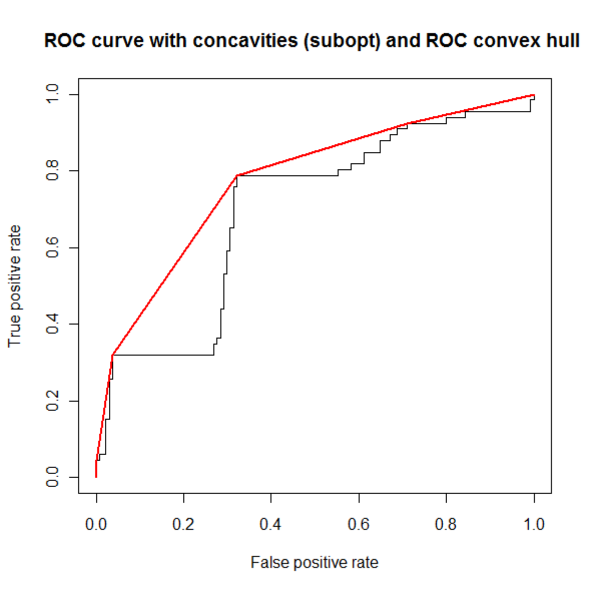
Accuracy vs Cutoff



Operating Conditions

- $p(+)$ and $p(-)$
- $p(+)$ in training set
- $p(+)$ in test set
- $p(+)$ in deployment

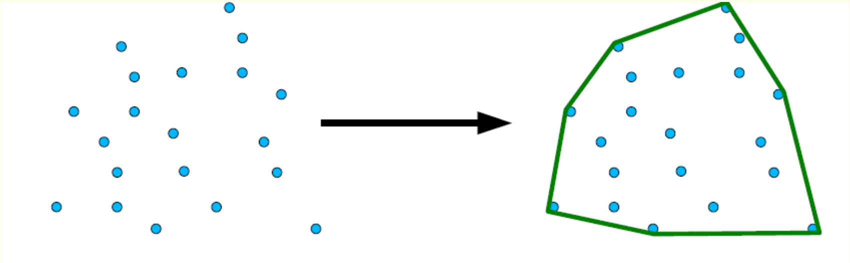
ROC and Convex hull



Convex Hull

The convex hull of a set of Q points is the smallest convex polygon P for which each point Q is either on the boundary of P or in its interior

Convex hull



Convex Hull again

- All points on the Convex Hull dominate
- We are looking for the point on the Convex Hull where the slope of the tangent = p(-)/p(+)
- Difficult to imagine slopes of tangents
- Different way of looking at it

A new space

We are going to plot Error vs p(+)

$$
\begin{aligned}
Err &= 1 - Acc \\
&= 1 - [p(+) * TPR + p(-) * (1 - FPR)] \\
&= p(+) * FNR + p(-) * FPR \\
&= (FNR - FPR) * p(+) + FPR
\end{aligned}
$$

New space Y-axis = Error rate; X-axis =p(+)

A new space

- Points in ROC space are mapped onto lines in Error vs p($+$) space
- Allows see how error depends on p($+$)
- How model will behave under various prior probabilities

Mapping two points onto a line

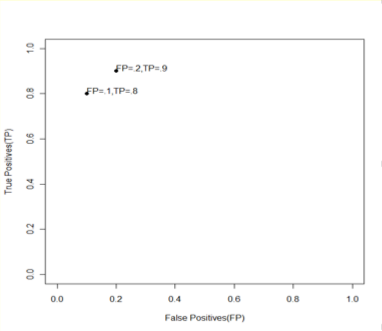
- FPR=0.1; TPR=0.8
- Easiest thing thing to do is see what happens when p(+) = 1 and p(+) =0
- p(+)=0 :
- Err = (FNR-FPR)*p(+)+FPR = FPR
- p(+) =1:
- Err = (FNR-FPR)*p(+)+FPR = FNR=1-TPR

For point FPR=0.1;TPR=0.8

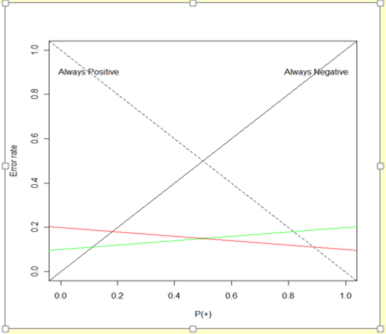
- p(+)=1; Err=1-TPR= 1-0.8=0.2
- p(+)=0: Err= FPR=0.1
- We can now draw line corresponding to this point

2 spaces

ROC space



"New Space"



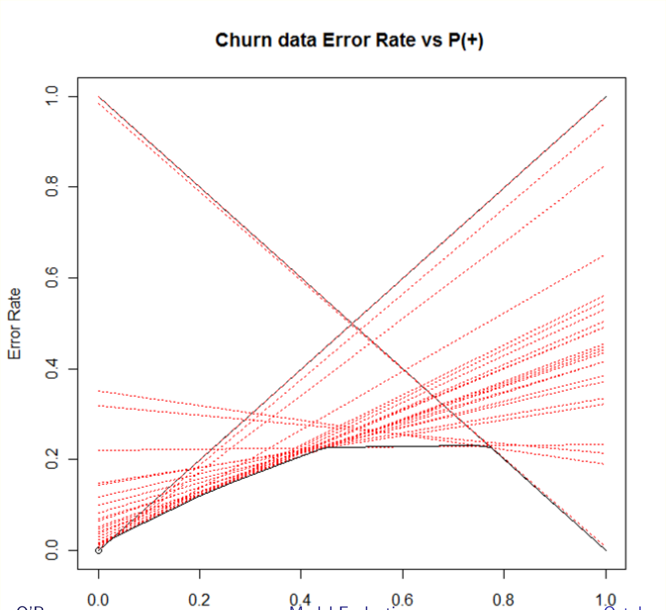
For point FPR=0;TPR=0

- Corresponds to Cutoff of 1; Always negative rule
- p(+)=1; Err=1-TPR= 1-0=1
- p(+)=0: Err= FPR=0
- We can now draw line

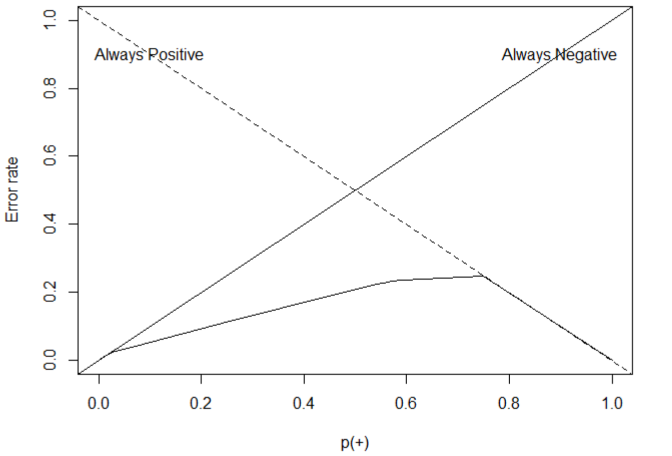
For point FPR=1;TPR=1

- Corresponds to Cutoff of 0; Always positive rule
- $p(+)=1$; Err=1-TPR= 1-1=0
- $p(+)=0$: Err= FPR=1
- We can now draw line

Churn data plot



Churn data plot



Incorporating Costs

	Predicted	
Actual	+	-
+	TP $C(+\|+)$	FN $C(+\|-)$
-	FP $C(-\|+)$	TN $C(-\|-)$

$C(+|-)$: Cost of misclassifying a + as a -

$C(-|+)$: Cost of misclassifying a - as +

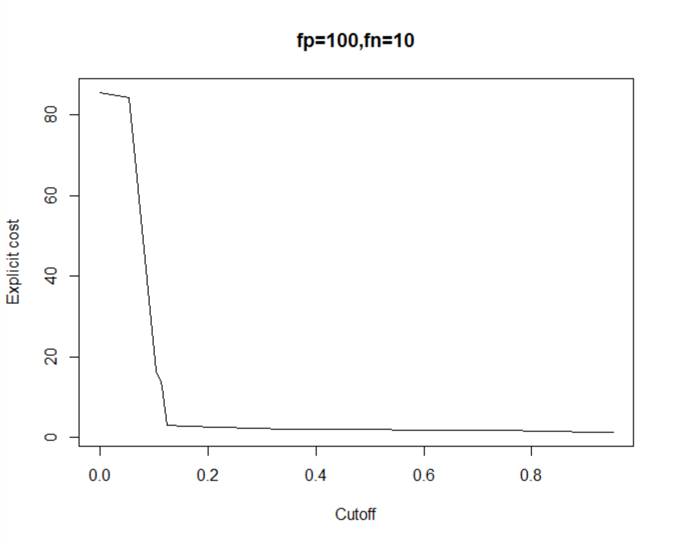
What do we now

- Total Cost = TP*$C(+|+)$+FN*$C(+|-)$+FP*$C(-|+)$ +TN*$C(-|-)$
- Want to minimise cost
- Assume $C(+|+)$ and $C(-|-)$ are 0
- Cost = FN*$C(+|-)$+FP*$C(-|+)$
- Sometimes written as Cost = (1-TP)*$C(+|-)$+FP*$C(-|+)$
- Have to estimate these costs

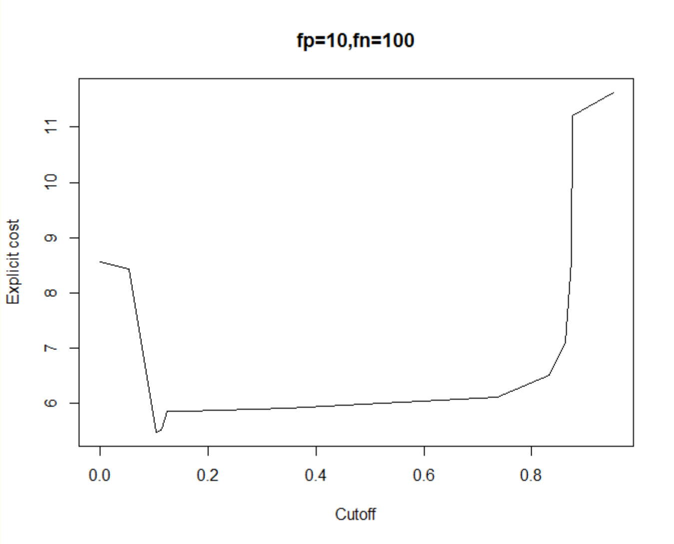
Difficulties

- Chunky model of cost
- Difficult to estimate the costs
- Ratio of costs are important
- Objective changes
- Many more complicated models

Costs and cutoff



Costs and cutoff



Costs and ROC curves

- Including priors
- Cost $=$ p($+$)FN*$C(+|-)$+p(-)FP*$C(-|+)$
- Cost $=$p($+$)(1-TP)*$C(+|-)$+p(-)FP*$C(-|+)$
- Define our performance lines in terms of costs and priors
- Slope now is $\frac{TP_1 - TP_2}{FP_1 - FP_2} = \frac{p(-)(C+|-)}{p(+)C(-|+)}$

How do we look at Costs and priors and ROC curve

- Map results into new space like before
- Define axes differently
- Examine model in this new space
- Tells us how models behave under various combinations of priors and costs

Looking at costs again

- Cost $=$ p($+$)FN*$C(+|-)$+p(-)FP*$C(-|+)$
- Max value for Cost when all missclassified;
- FN$=$1 and FP$=$1
- Then Cost $=$ p($+$)$C(+|-)$+p(-)$C(-|+)$
- Use to normalise cost to get

$$NE(C) = \frac{(1-TP)p(+)C(+|-) + FPp(-)C(-|+)}{p(+)C(+|-) + p(-)C(-|+}$$

Looking at costs again

- Define a new function of priors and costs

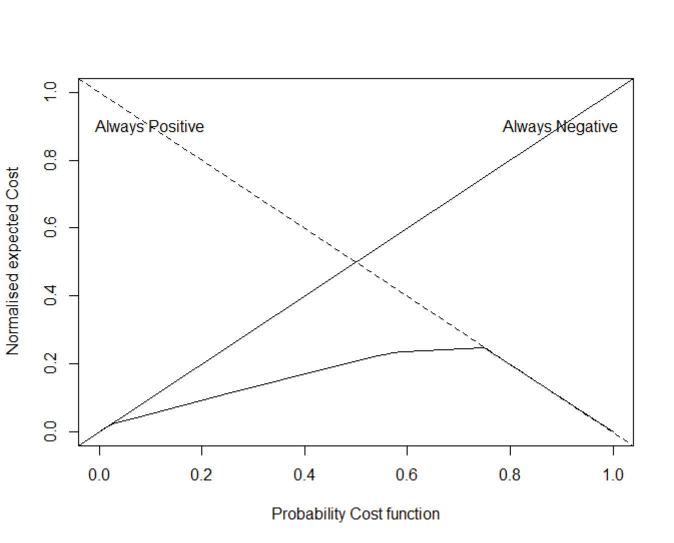
$$PCF(+) = \frac{p(+)C(+|-)}{p(+)C(-|+) + p(-)C(+|-)}$$

- Probability Cost function
- This ranges from 0 to 1
- For equal costs PCF$(+)$ =p$(+)$
- Plot PCF$(+)$ vs NE(C)

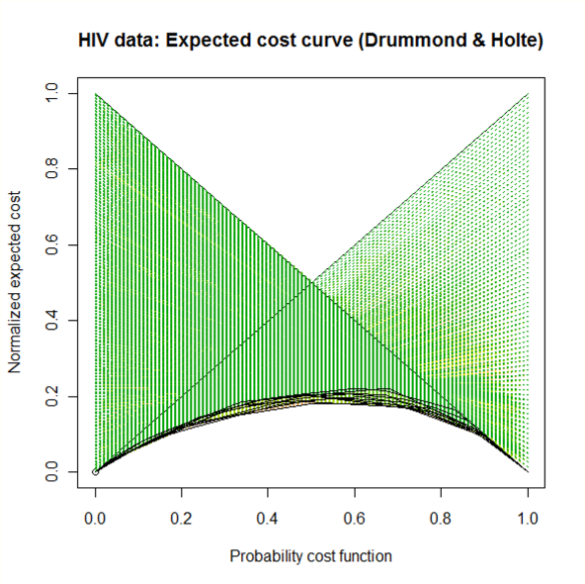
And looking at this closer

- NE(C) can be written as (FN-FP)*PCF(+) + FP
- Before we had
- ERR= (FN-FP)*p(+) + FP
- What has changed is interpretation of the axes.

PCF vs Normalised cost



A very pretty picture



How do we use this?

- Draw chart as before
- Calculate PCF(+) for various values of p(+), $C(+|-)$, $C(-|+)$
- See where they fall in chart
- Are they better than a trivial model?

Calculating values for PCF

$p(+)$	$C(+\|-)$	$C(-\|+)$	$PCF(+)$
0.2	20	10	0.33
0.2	10	20	0.11
0.2	50	10	0.56
0.8	50	10	0.95
0.8	10	50	0.44
0.5	10	50	0.17

A summary

- Confusion tables for each cutoff
- Calculate sensitivity and specificity for each table
- Draw ROC curve
- Calculate AUC + SE
- Investigate the effect of priors and costs on the result if appropriate