



Opinionated Lessons in Statistics

by Bill Press

#50 Binary Classifiers

A (binary) classifier classifies data points as + or -

If we also know the true classification, the performance of the classifier is a 2x2 contingency table, in this application usually called a confusion matrix.

		actual	
		+	-
classifier	+	TP good!	FP
	-	FN	TN good!

bad! (Type I error)

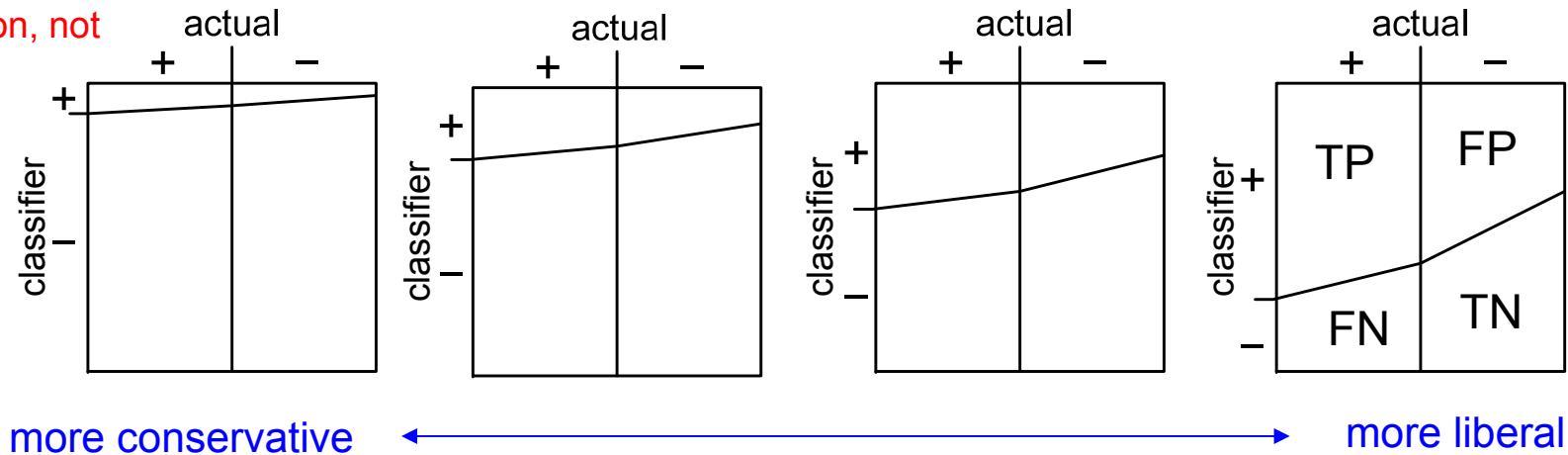
bad! (Type II error)

As we saw, this kind of table has many other uses: treatment vs. outcome, clinical test vs. diagnosis, etc.

Earlier we were looking at statistically “weak” contingency tables and trying to decide if they were significant. Here we’re interested in the strength of the signal; the (high) significance is a given.

Most classifiers have a “knob” or threshold that you can adjust: How certain do they have to be before they classify a “+”? To get more TP’s, you have to let in some FP’s!

Cartoon, not
literal:



Notice there is just one free parameter, think of it as TP, since

$$FP(TP) = [\text{given by algorithm}]$$

$$TP + FN = P \text{ (fixed number of actual positives, column marginal)}$$

$$FP + TN = N \text{ (fixed number of actual negatives, column marginal)}$$

So all scalar measures of performance are functions of one free parameter (i.e., curves).

And the points on any such curve are in 1-to-1 correspondence with those on any other such curve.

If you ranked some classifiers by how good they are, you might get a different rankings at different points on the scale.

On the other hand, one classifier might dominate another at all points on the scale.

Terminology used to measure the performance of classifiers

Different combinations of ratios have been given various names.

All vary between 0 and 1.

A performance curve picks one as the independent variable and looks at another as the dependent variable.

		actual	-
classifier	+	TP	FP Type I error
	-	FN	TN Type II error

		actual	-
classifier	+	TP	FP
	-	FN	TN

accuracy (ACC)

		actual	-
classifier	+	TP	FP
	-	FN	TN

true pos rate (TPR)
 \equiv sensitivity
 \equiv recall

		actual	-
classifier	+	TP	FP
	-	FN	TN

precision-recall curve
pos. predictive value (PPV)
 \equiv precision

\Updownarrow "one minus"

		actual	-
classifier	+	TP	FP
	-	FN	TN

neg. predictive value (NPV)

		actual	-
classifier	+	TP	FP "one minus"
	-	FN	TN

specificity (SPC)

\Updownarrow "one minus"

		actual	-
classifier	+	TP	FP
	-	FN	TN

false pos rate (FPR)
ROC curve

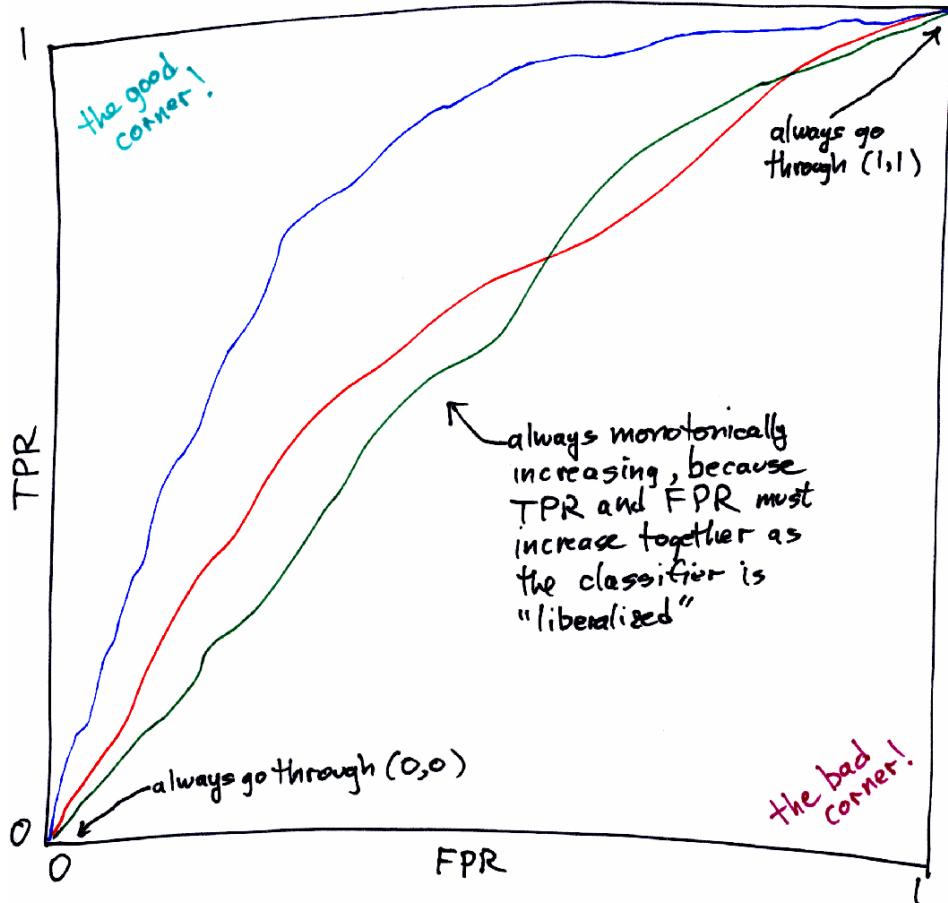
		actual	-
classifier	+	TP	FP
	-	FN	TN

false discovery rate (FDR)

Dark color is numerator, dark and light color is denominator.

Blue parameters: 1 is good. Red: 0 is good.

ROC (“Receiver Operating Characteristic”) curves plot TPR vs. FPR as the classifier goes from “conservative” to “liberal”



blue dominates red and green
neither red nor green dominate the other

actual	+	-
	TP	FP
classifier +	FN	TN
-		

true pos rate (TPR)
 \equiv sensitivity
 \equiv recall

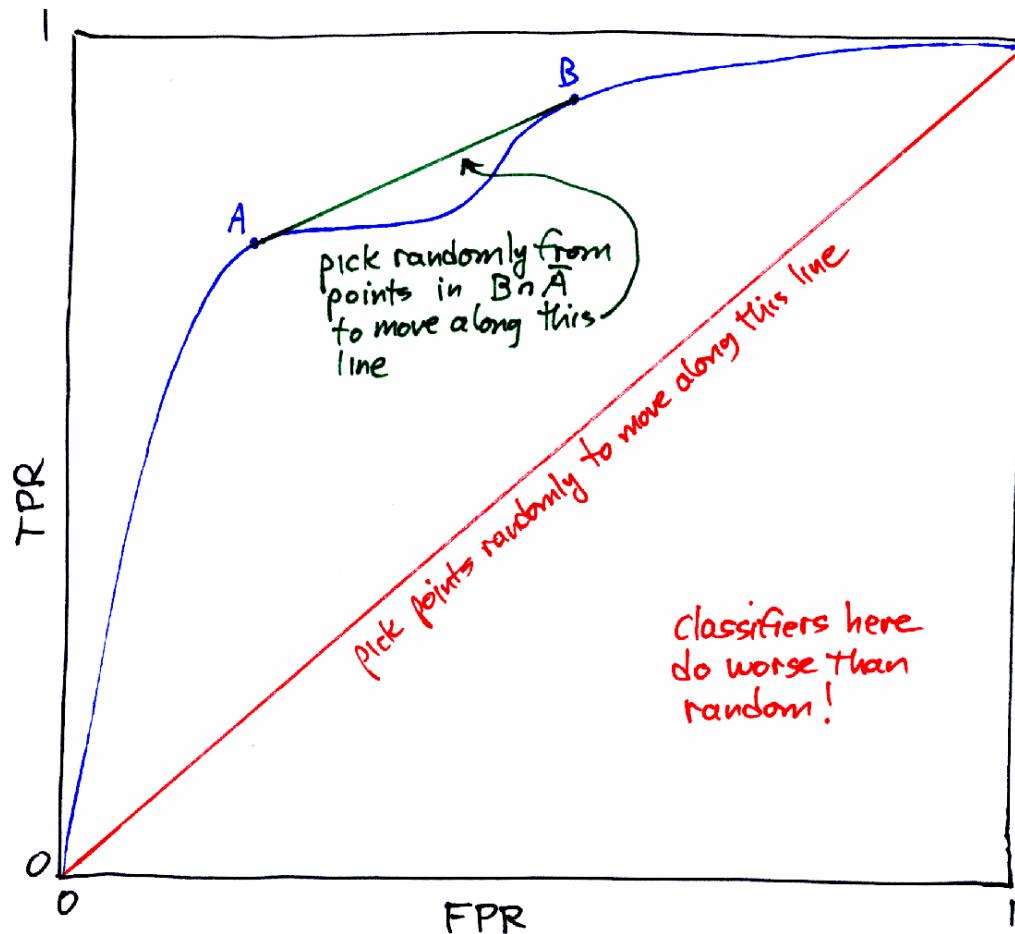
actual	+	-
	TP	FP
classifier +	FN	TN
-		

false pos rate (FPR)

ROC curve

You could get the best of the red and green curves by making a hybrid or “Frankenstein” classifier that switches between strategies at the cross-over points.

ROC curves can always be “upgraded” to their convex hull by replacing any concave portions by a random sample



List points classified as + by B but not A.

Start up the curve to A.

When you reach A, start adding a fraction of them (increasing from 0 to 1) randomly, until you reach B.

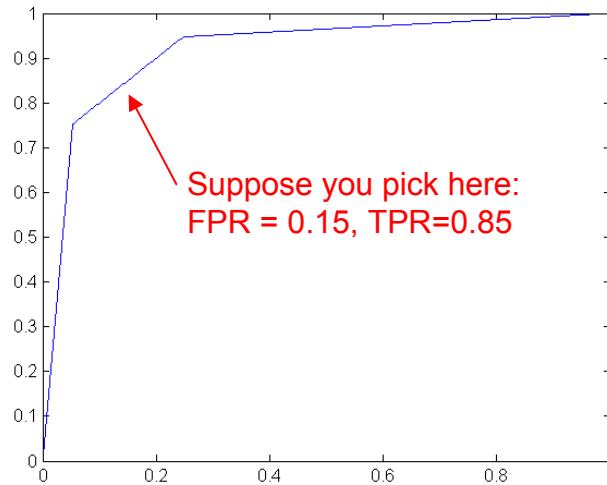
Continue on the curve from B.

Of course to measure the ROC curve at all, you have to have known “training” or “ground truth” data. You use that data to estimate the points A and B, then create the convex (upgraded) classifier.

Since ROC curves don't explicitly show any dependence on the constant P/N (ratio of actual + to - in the sample) they can be misleading if you care about, say, FP versus TP (or any other cross-column comparison).

Suppose you have a test for Alzheimer's whose false positive rate can be varied from 5% to 25% as the false negative rate varies from 25% to 5% (suppose linear dependences on both):

```
I.am = (0: 0.01: 1);
fpr = .05 + 0.2 * I.am;
tpr = 1 - (.05 + 0.2*(1-I.am));
fpr(1) = 0;
fpr(end) = 1;
tpr(1) = 0;
tpr(end) = 1;
plot(fpr, tpr)
```



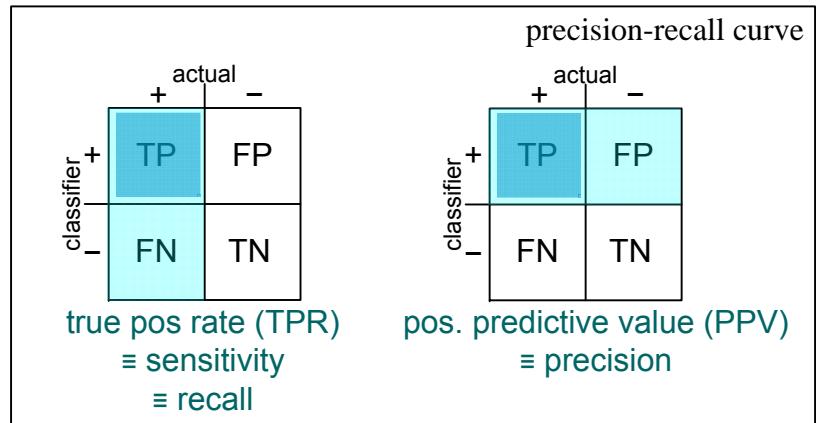
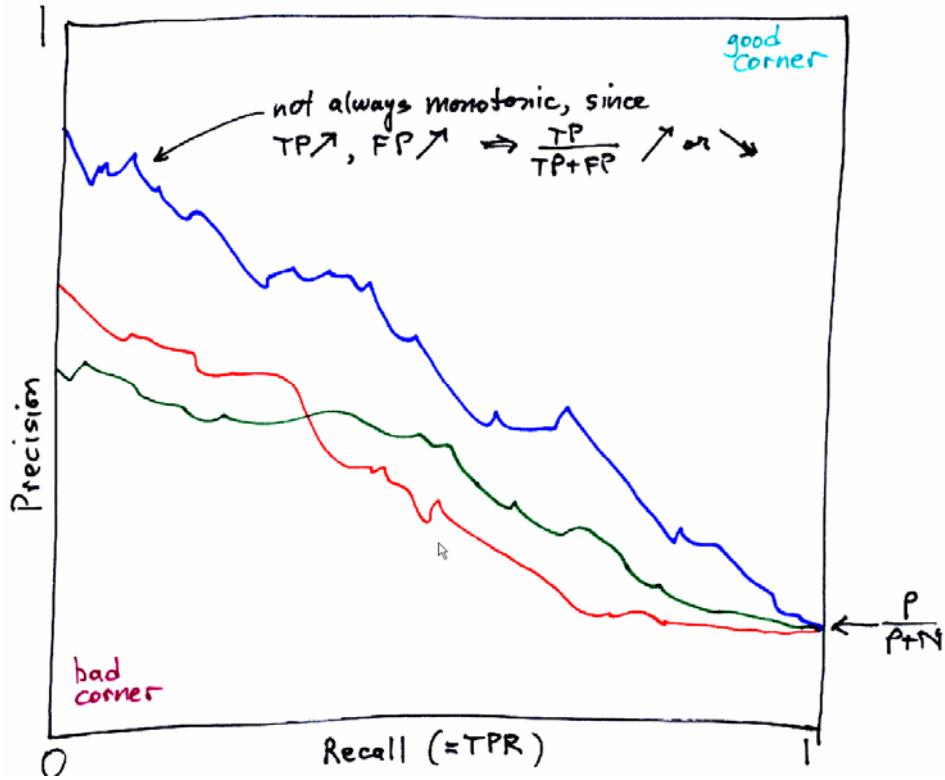
Now suppose you try the test on a population of 10,000 people, 1% of whom actually are Alzheimer's positive:

		actual
		+
classifier	+	85
	-	1485
-	+	15
	-	8415

FP swamps TP by ~17:1. You'll be telling 17 people that they might have Alzheimer's for every one who actually does. It is unlikely that your test will be used.

In a case like this, ROC, while correct, somewhat misses the point.

Precision-Recall curves overcome this issue by comparing TP with FN and FP

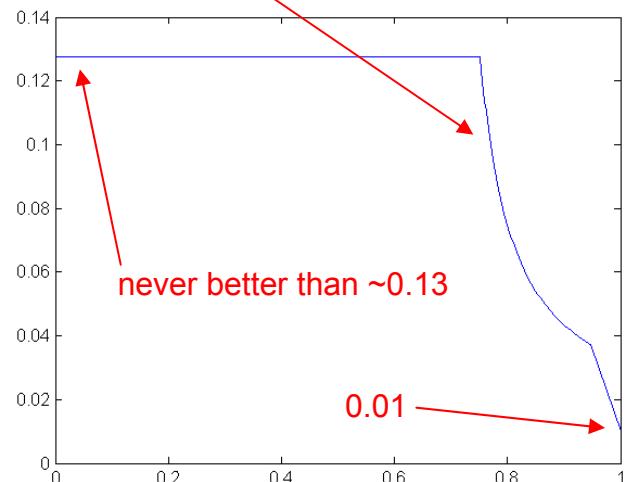


Continue our toy example:

note that P and N now enter

```
prec = tpr*100. / (tpr*100+fpr*9900);
prec(1) = prec(2); % fix up 0/0
reca = tpr;
plot(reca, prec)
```

By the way, this shape "cliff" is what the ROC convexity constraint looks like in a Precision-Recall plot. It's not very intuitive.



For fixed marginals P,N the points on the ROC curve are in 1-to-1 correspondence with the points on the Precision-Recall curve.

That is, both display the same information. You can go back and forth.

$$\left. \begin{aligned} \text{pre} &= \frac{\text{TPR } P}{\text{TPR } P + \text{FPR } N} \\ \text{rec} &= \text{TPR} \\ \frac{\text{rec} (1 - \text{pre})}{\text{pre}} \frac{P}{N} &= \text{FPR} \end{aligned} \right\} \begin{array}{l} \text{TPR, FPR from pre, rec} \\ \text{pre, rec from TPR, FPR} \end{array}$$

It immediately follows that if one curve dominates another in ROC space, it also dominates in Precision-Recall space.

(Because a crossing in one implies a crossing in the other, by the above equations.)

But for curves that cross, the metrics in one space don't easily map to the other. For example, people sometimes use “area under the ROC curve”. This doesn't correspond to “area under the Precision-Recall curve”, or to anything simple.

One also sees used PPV and NPV
(more often as a pair of numbers than as a curve)

PPV: given a positive test, how often does the patient have the disease.

NPV: given a negative test, how often is the patient disease-free.

		actual	
		+	-
classifier	+	TP	FP
	-	FN	TN

pos. predictive value (PPV)
≡ precision

		actual	
		+	-
classifier	+	TP	FP
	-	FN	TN

neg. predictive value (NPV)

		actual	
		+	-
classifier	+	85	1485
	-	15	8415

$$\text{PPV} = 0.054$$

$$\text{NPV} = 0.998$$

So could a physician use this test “to rule out Alzheimers” in a case that presents with some symptoms?

No, because in the population of people who present, the ratio of the columns would be not nearly so extreme.

You have to be careful about asking exactly the question you want!

It's easy to get from PPV,NPV to ROC or vice versa. Or, for that matter, any other of the parameterizations. In Mathematica, for example:

```
In[1]:= eqs = {PPV == TP / (TP + FP), NPV == TN / (TN + FN), TP + FN == P, TN + FP == N}
```

$$\text{Out}[1]= \left\{ \text{PPV} == \frac{\text{TP}}{\text{FP} + \text{TP}}, \text{NPV} == \frac{\text{TN}}{\text{FN} + \text{TN}}, \text{FN} + \text{TP} == \text{P}, \text{FP} + \text{TN} == \text{N} \right\}$$

```
In[2]:= ans = FullSimplify[Solve[eqs, {TP, FP, TN, FN}]]
```

$$\text{Out}[2]= \left\{ \begin{aligned} \text{FN} &\rightarrow -\frac{(-1 + \text{NPV}) (\text{P} (-1 + \text{PPV}) + \text{N} \text{PPV})}{-1 + \text{NPV} + \text{PPV}}, \text{TP} \rightarrow \frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) \text{PPV}}{-1 + \text{NPV} + \text{PPV}}, \\ \text{FP} &\rightarrow -\frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) (-1 + \text{PPV})}{-1 + \text{NPV} + \text{PPV}}, \text{TN} \rightarrow \frac{\text{NPV} (\text{P} (-1 + \text{PPV}) + \text{N} \text{PPV})}{-1 + \text{NPV} + \text{PPV}} \end{aligned} \right\}$$

```
In[4]:= TPR = FullSimplify[TP / (TP + FN) /. ans]
```

$$\text{Out}[4]= \left\{ \frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) \text{PPV}}{\text{P} (-1 + \text{NPV} + \text{PPV})} \right\}$$

```
In[5]:= FPR = FullSimplify[FP / (FP + TN) /. ans]
```

$$\text{Out}[5]= \left\{ -\frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) (-1 + \text{PPV})}{\text{N} (-1 + \text{NPV} + \text{PPV})} \right\}$$

```
In[8]:= eqs2 = {tpr == \left( \frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) \text{PPV}}{\text{P} (-1 + \text{NPV} + \text{PPV})} \right), fpr == \left( -\frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) (-1 + \text{PPV})}{\text{N} (-1 + \text{NPV} + \text{PPV})} \right)}
```

$$\text{Out}[8]= \left\{ \text{tpr} == \frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) \text{PPV}}{\text{P} (-1 + \text{NPV} + \text{PPV})}, \text{fpr} == -\frac{(\text{N} (-1 + \text{NPV}) + \text{NPV} \text{P}) (-1 + \text{PPV})}{\text{N} (-1 + \text{NPV} + \text{PPV})} \right\}$$

```
In[9]:= FullSimplify[Solve[eqs2, {NPV, PPV}]]
```

$$\text{Out}[9]= \left\{ \left\{ \text{NPV} \rightarrow \frac{(-1 + \text{fpr}) \text{N}}{(-1 + \text{fpr}) \text{N} + \text{P} (-1 + \text{tpr})}, \text{PPV} \rightarrow \frac{\text{P} \text{tpr}}{\text{fpr} \text{N} + \text{P} \text{tpr}} \right\} \right\}$$