Digital Transformation of Healthcare

Evaluating Predictions

Michoel Snow, MD PhD, Glen Ferguson, PhD

Center for Health Data Innovations

Objectives

After this lecture students will be able to

- Distinguish between classification and regression metrics
- Compare and contrast the use of different metrics to evaluate predictions
-

Metrics for Evaluation of Classification Models

Confusion Matrix

predicted outcome	actual outcome: p	actual outcome: n	
p′	TP	FP	$\text{PPV} = \frac{TP}{TP + FP}$
n′	FN	TN	$\text{NPV} = \frac{TN}{FN + TN}$
	$\text{Sens} = \frac{TP}{TP + FN}$	$\text{Spec} = \frac{TN}{FP + TN}$	$\text{Acc} = \frac{TP+TN}{TP+TN+FP+FN}$

Confusion Matrix

predicted outcome	actual outcome: **p**	actual outcome: **n**	
p′	TP	FP	$\text{PPV} = p(\mathbf{p} \mid \mathbf{p}')$
n′	FN	TN	$\text{NPV} = p(\mathbf{n} \mid \mathbf{n}')$
	$\text{Sens} = p(\mathbf{p}' \mid \mathbf{p})$	$\text{Spec} = p(\mathbf{n}' \mid \mathbf{n})$	$\text{Acc} = p(TP + TN)$

Confusion Matrix

		actual outcome **p**	actual outcome **n**	
predicted outcome	$\mathbf{p}'$	TP	FP	$\text{PPV} = p(\mathbf{p} \mid \mathbf{p}')$
	$\mathbf{n}'$	FN	TN	$\text{NPV} = p(\mathbf{n} \mid \mathbf{n}')$
		$\text{Sens} = p(\mathbf{p}' \mid \mathbf{p})$	$\text{Spec} = p(\mathbf{n}' \mid \mathbf{n})$	$\text{Acc} = p(TP + TN)$

Parameter	Interpretation	Inappropriate for
Accuracy	Overall proximity of test to reality	Imbalanced sample sizes
Sensitivity		
Specificity		
PPV		
NPV		

Confusion Matrix

		actual outcome **p**	actual outcome **n**	
predicted outcome	$\mathbf{p}'$	TP	FP	$\text{PPV} = p(\mathbf{p} \mid \mathbf{p}')$
	$\mathbf{n}'$	FN	TN	$\text{NPV} = p(\mathbf{n} \mid \mathbf{n}')$
		$\text{Sens} = p(\mathbf{p}' \mid \mathbf{p})$	$\text{Spec} = p(\mathbf{n}' \mid \mathbf{n})$	$\text{Acc} = p(TP + TN)$

Parameter	Interpretation	Inappropriate for
Accuracy	Overall proximity of test to reality	Imbalanced sample sizes
Sensitivity	Chance of a false negative	Expensive testing/Mild disease
Specificity	Chance of a false positive	Cheap testing/Severe disease
PPV	Sensitivity diagnostic utility	Very high prevalence
NPV	Specificity diagnostic utility	Very low prevalence

Combined Statistics

Function of	Metric	Formula
Sensitivity, Specificity	Positive Likelihood Ratio/ROC	$\frac{sensitivity}{1 - specificity}$
Sensitivity, Specificity	Negative Likelihood Ratio	$\frac{1 - sensitivity}{specificity}$
Sensitivity, PPV	F1 score	$\frac{2}{\frac{1}{sensitivity} + \frac{1}{PPV}}$
TP, TN, FP, FN	Matthews correlation coefficient	$\frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

Likelihood Ratios

Does a test result change the probability that a person has a certain condition?

$$LR+ = \frac{sensitivity}{1 - specificity} = \frac{P(T+ \mid D+)}{P(T+ \mid D-)}$$

$$LR- = \frac{1 - sensitivity}{specificity} = \frac{P(T- \mid D+)}{P(T- \mid D-)}$$

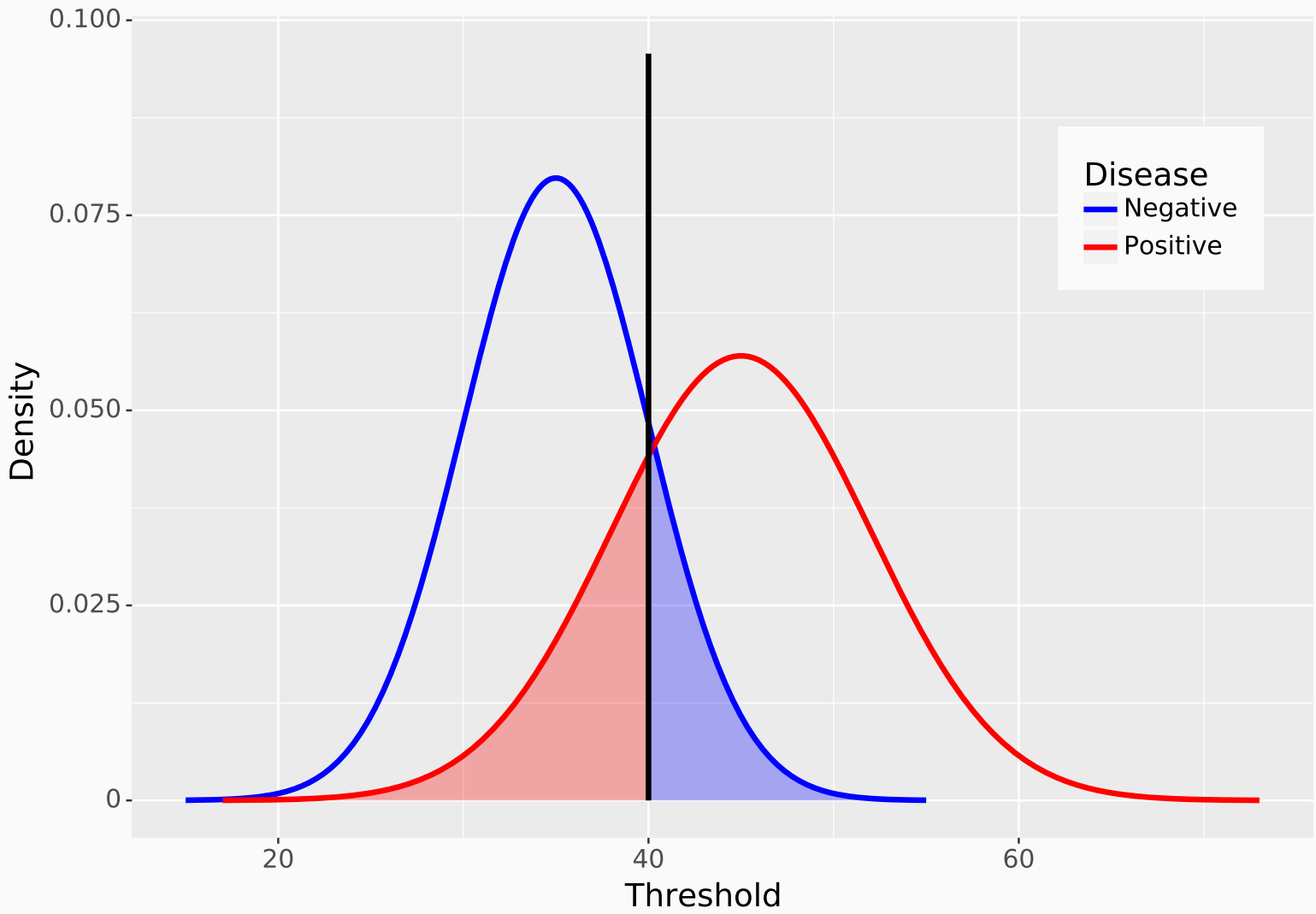
Likelihood Ratios

Likelihood Ratio	Approximate Change in Probability(%)
0.1	-45
0.2	-30
0.5	-15
1	0
2	+15
5	+30
10	+45

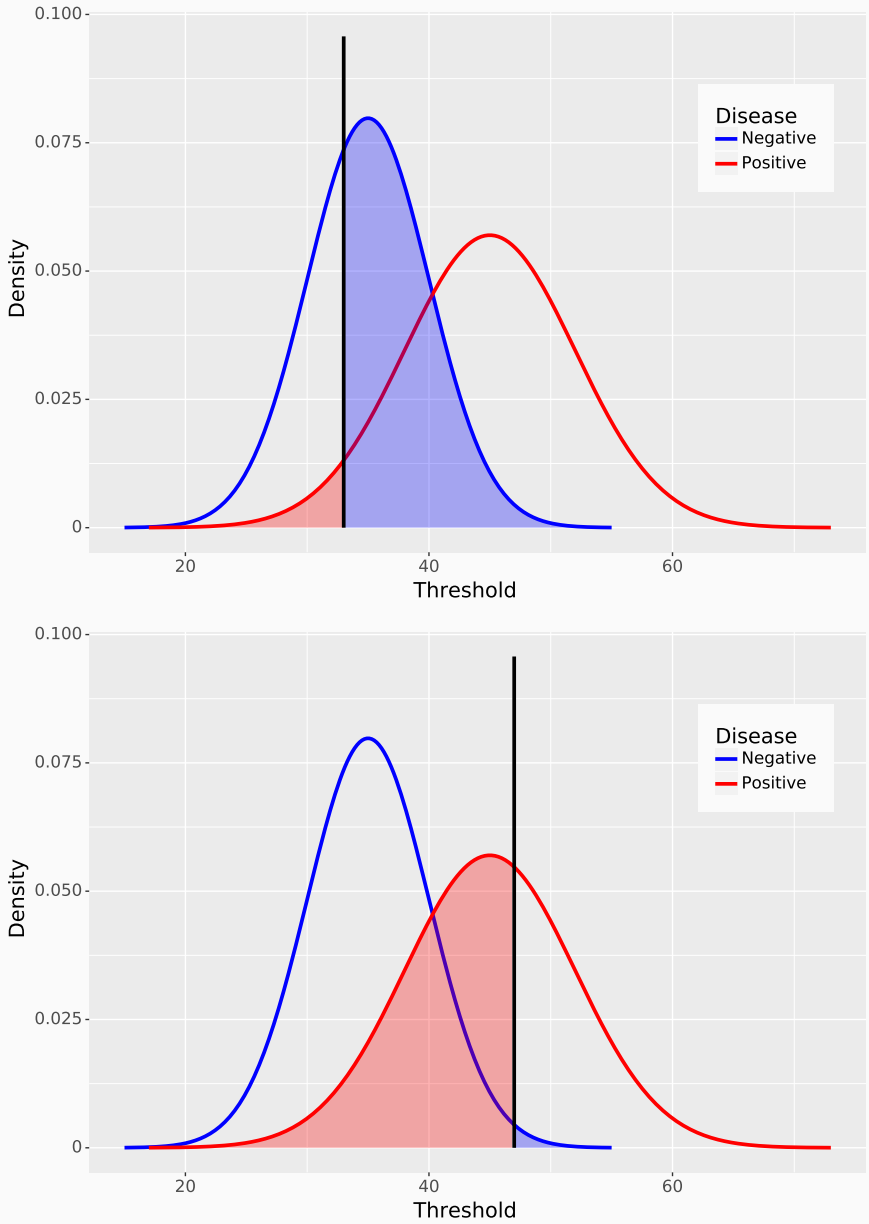
$$\text{Change in post test probability} \approx 0.2 \times \ln LR \ ^{1}$$

[1] McGee, Steven. "Simplifying likelihood ratios." Journal of general internal medicine 17.8 (2002): 647-650. APA

Hypothesis Testing



Hypothesis Testing



Receiver Operating Characteristic

How can I evaluate a binary classifier over its range of possible discrimination thresholds?