Measuring Performance in Supervised Learning 3

-Applied Analytics-

APM Chapter 11.3 & 11.4

Lecturer: Darren Homrighausen, PhD

Preamble:

- Discuss the probability threshold for turning $\hat{p}$ into $\hat{Y}$
- Cover the receiver operation characteristic

The probability threshold

We mentioned how Y_* is formed by choosing the class that maximizes p_ℓ

Hence, it is typical to make a classifier that mimics Y_* and form $\hat{Y}$ that maximizes $\hat{p}_\ell$

This corresponds to comparing the $\hat{p}_\ell$ to threshold $1/C$

However, there are cases in which a different threshold is natural

The probability threshold

Example: We could train a classifier to classify whether a plane needs to be serviced

		Truth	
		Serviced	Not serviced
Our Preds.	Serviced	Plane gets fixed	plane could have been flying
	Not serviced	Potential crash	plane flies

Here, the economic costs of not having a plane available when it could have been don't compare to flying an unfit plane

One way to incorporate these ideas is via adjusting the threshold

Sensitivity and specificity

In this example, the 'interesting' case is the plane needs to be serviced

Hence,

- sensitivity is $\text{prob}(\hat{Y} = \text{serviced} | Y = \text{serviced})$
  (and the $\widehat{\text{prob}}(\hat{Y} = \text{serviced} | Y = \text{serviced}) = TP/P$)
- specificity is $\text{prob}(\hat{Y} = \text{not serviced} | Y = \text{not serviced})$
  (and the $\widehat{\text{prob}}(\hat{Y} = \text{not serviced} | Y = \text{not serviced}) = TN/N$)

The probability of the critical error in this case is 1 - sensitivity

Using the 0.5 threshold might result in a specificity of .87 and sensitivity of .34

Question: If we have C_1 = 'serviced' and adjust the threshold to $\hat{p}_1 > .8$, then will the sensitivity go up or down?

Sensitivity and specificity

In this example, the 'interesting' case is the plane needs to be serviced

Hence,

- sensitivity is $\text{prob}(\hat{Y} = \text{serviced} | Y = \text{serviced})$
  (and the $\widehat{\text{prob}}(\hat{Y} = \text{serviced} | Y = \text{serviced}) = TP/P$)
- specificity is $\text{prob}(\hat{Y} = \text{not serviced} | Y = \text{not serviced})$
  (and the $\widehat{\text{prob}}(\hat{Y} = \text{not serviced} | Y = \text{not serviced}) = TN/N$)

The probability of the critical error in this case is 1 - sensitivity

Using the 0.5 threshold might result in a specificity of .87 and sensitivity of .34

Question: If we have C_1 = 'serviced' and adjust the threshold to $\hat{p}_1 > .8$, then will the sensitivity go up or down?
Down!

Receiver operating characteristic

We can therefore adjust the sensitivity and specificity by adjusting the threshold

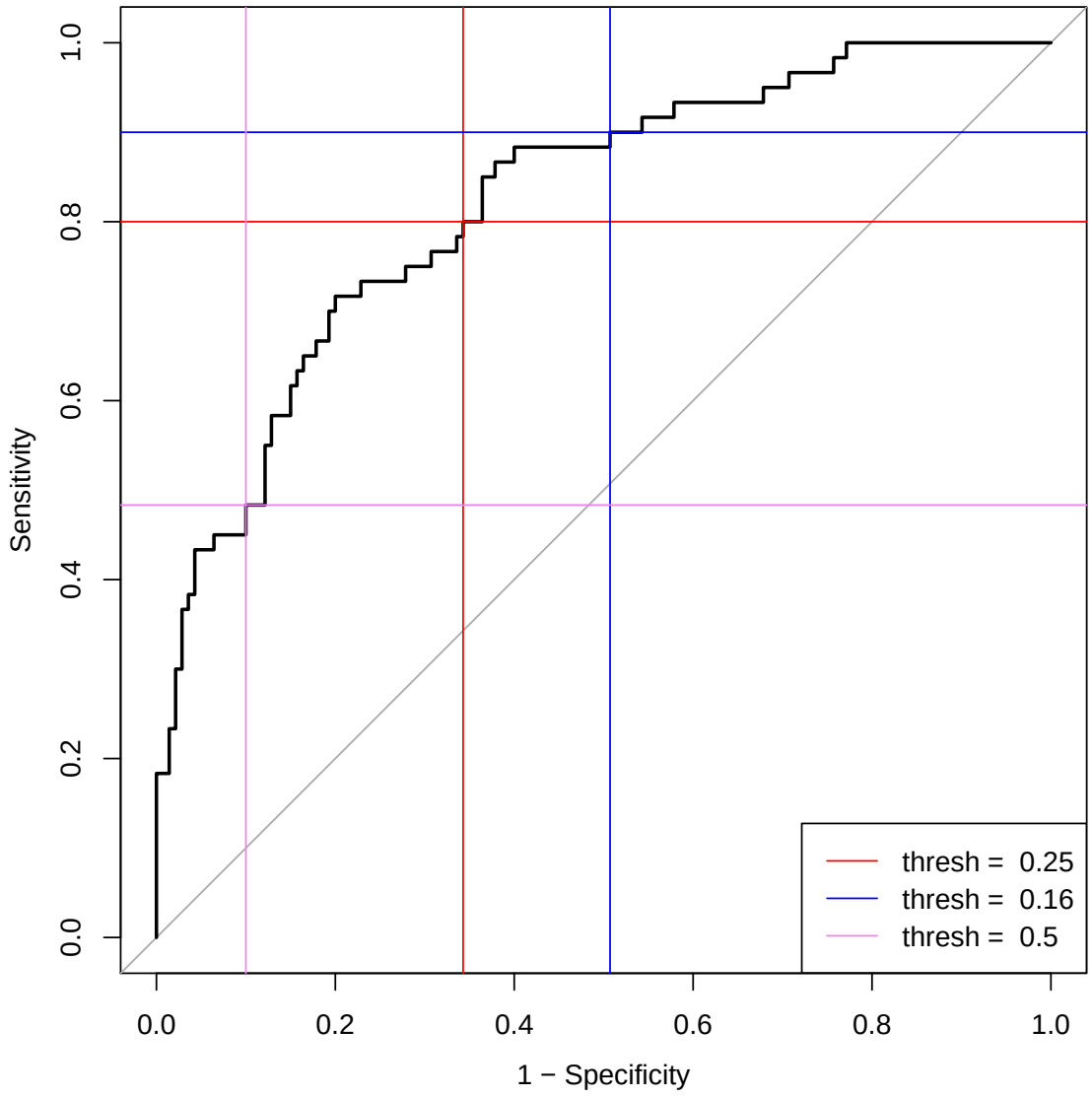
In face, this threshold is another example of a tuning parameter

We will get a (potentially) new classifier for any value of the threshold from (0,1)

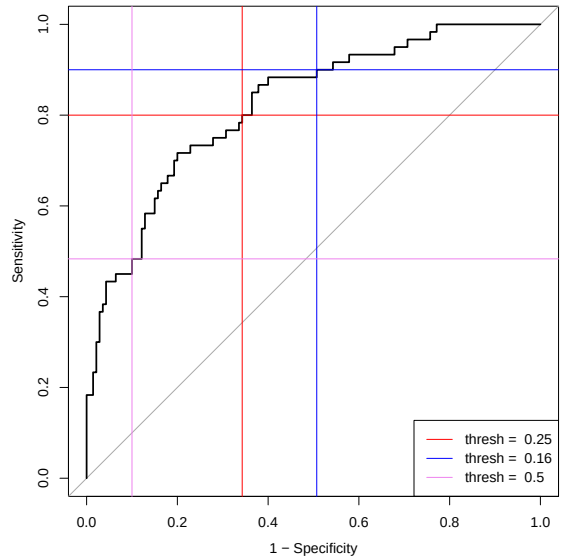
The receiver operating characteristic (ROC) plots three things:

- The sensitivity/recall
- 1-specificity (false positive)
- the threshold

Receiver operating characteristic



Receiver operating characteristic



The ROC plot can be used in a few different ways

- Perhaps we state that we will not accept a classifier unless it has at least a sensitivity of .8 on test data
  $\rightarrow$ set the threshold at least 0.25
- Alternatively, we can get a quantitative measurement about a classifier averaged across all possible values of the threshold
  $\rightarrow$ get the area under the ROC curve

Area under the ROC curve

The area under the ROC curve give a summary of the plot and is called AUC

$(0 \leq \mathrm{AUC} \leq 1)$

The interpretation: a procedure with larger AUC is better and AUC ≈ 1 is best

The ROC curve, like sensitivity and specificity, are only defined for $C = 2$

 My opinion: I find the AUC worthless. But it is used a lot in practice. 

Postamble:

- Discuss the probability threshold for turning $\hat{p}$ into $\hat{Y}$

  (Varying the threshold, we can make classifier have better sensitivity or specificity properties. However, increasing one almost always decreases the other)
- Cover the receiver operation characteristic

  (The ROC curve gives a visual summary of the sensitivity and (one minus the) specificity as a function of the threshold)