Prediction

- Project 1, Part 1 feedback sent
- HW 6 released. Due in two weeks
- Exam 1 next Friday
- Today:
  - wrap up prediction
  - work time

Last time: Class activity

As threshold ↑, sensitivity ↓
Specificity ↑

As threshold ↓, sensitivity ↑
Specificity ↓

Changing thresholds:

```
table(Prediction = m1$fitted.values > 0.3,
      Truth = sba$Default)
```

```
##            Truth
## Prediction FALSE TRUE
##      FALSE  3524  351
##      TRUE    565  551
```

```
table(Prediction = m1$fitted.values > 0.7,
      Truth = sba$Default)
```

```
##            Truth
## Prediction FALSE TRUE
##      FALSE  4089  902
```

Changing thresholds

> How can I assess prediction performance across many different thresholds?

- Look at many thresholds e.g. $0, 0.01, 0.02, \ldots, 0.99, 1$
- For each threshold, calculate metrics like sensitivity and specificity
- Plot sensitivity vs. 1-specificity

  $\Rightarrow$ Receiver operating characteristic (ROC) curve

plot trade-off between sensitivity and specificity

ROC curve

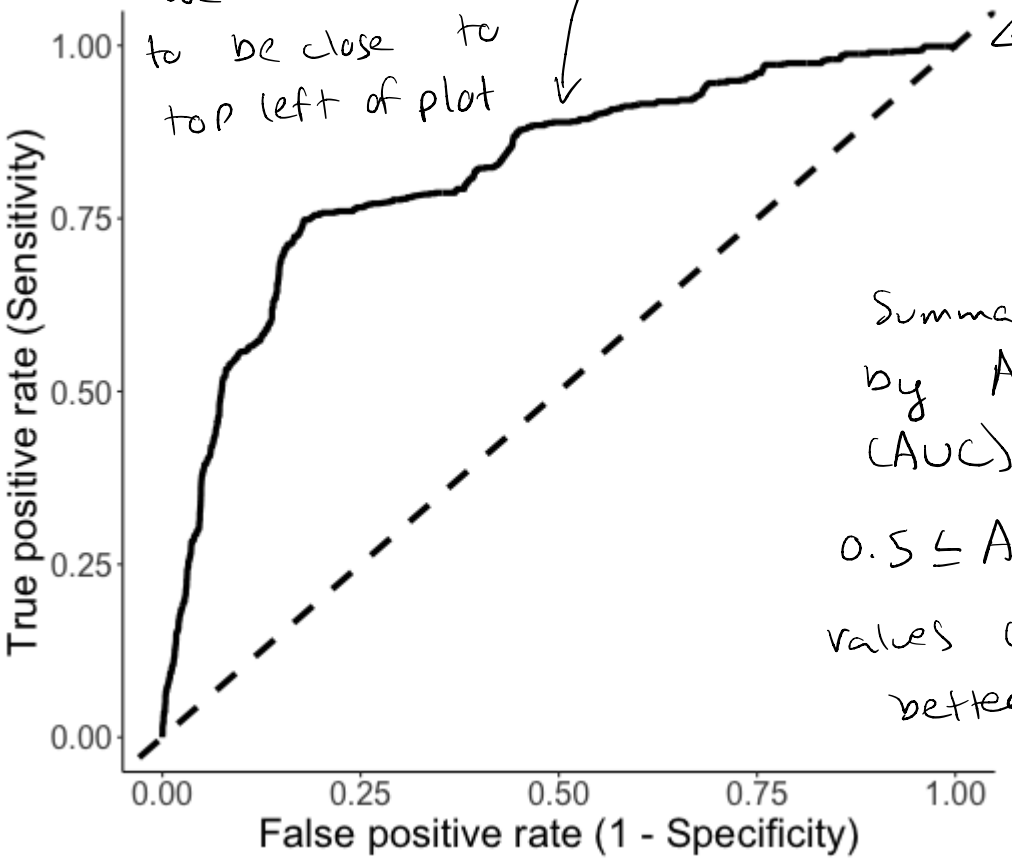
each point corresponds to one possible threshold

we want ROC curve to be close to top left of plot

curve will never fall below diagonal line

In this example:

AUC ≈ 0.82



Summarize ROC curve by Area Under Curve (AUC)

$0.5 \leq AUC \leq 1$

values closer to 1 are better

Homework 6

- Reading the dengue research paper
- Reproducing the paper's results
- Practice with model selection and assessing prediction performance