Tuning LDA

Confusion Matrix: Fit Statistics

- Example (see textbook): predicting loan defaults with LDA:

$$Error\ Rate = \frac{Incorrect}{Total} = \frac{23 + 252}{10{,}000} = 0.028$$

$$Sensitivity = \frac{Correct\ Defaults}{(+)\ Predictions} = \frac{81}{333} = 0.243$$

$$Specificity = \frac{Correct\ (-)}{(-)\ Predictions} = \frac{9{,}644}{9{,}667} = 0.997$$

Predicted Default	Actual Default		Total
	No	**Yes**	
No	9,644	252	**9,896**
Yes	23	81	**104**
Total	**9,667**	**333**	**10,000**

- Interestingly, this model has a **low error rate** of 2.8%; but we know **a priori** that **3.3%** (333/10,000) of the clients default, so we could predict that **no one** will **default** and be correct **96.7%** of the time → The model is not a big help in this respect.
- Similarly, the models does a **poor job** (worse than flipping a coin) at **predicting** actual **defaults**
- In contrast, it does a very **nice job** at predicting **no-defaults**.

Tuning LDA: The Threshold

- LDA uses a **"classifier threshold"** $\lambda = Pr(Default = Yes) > 0.5$ by default. So, only loans with more than a **50%** chance of defaulting are classified as expected defaults. But what if we change this threshold to a more conservative **20%**? See what happens:

$$Error\ Rate = \frac{Incorrect}{Total} = \frac{253 + 138}{10{,}000} = 0.039$$

$$Sensitivity = \frac{Correct\ Defaults}{(+)\ Predictions} = \frac{195}{333} = 0.586$$

$$Specificity = \frac{Correct\ (-)}{(-)\ Predictions} = \frac{9{,}432}{9{,}667} = 0.976$$

Predicted Default	Actual Default		Total
	No	Yes	
No	9,432	138	**9,896**
Yes	235	195	**104**
Total	**9,667**	**333**	**10,000**

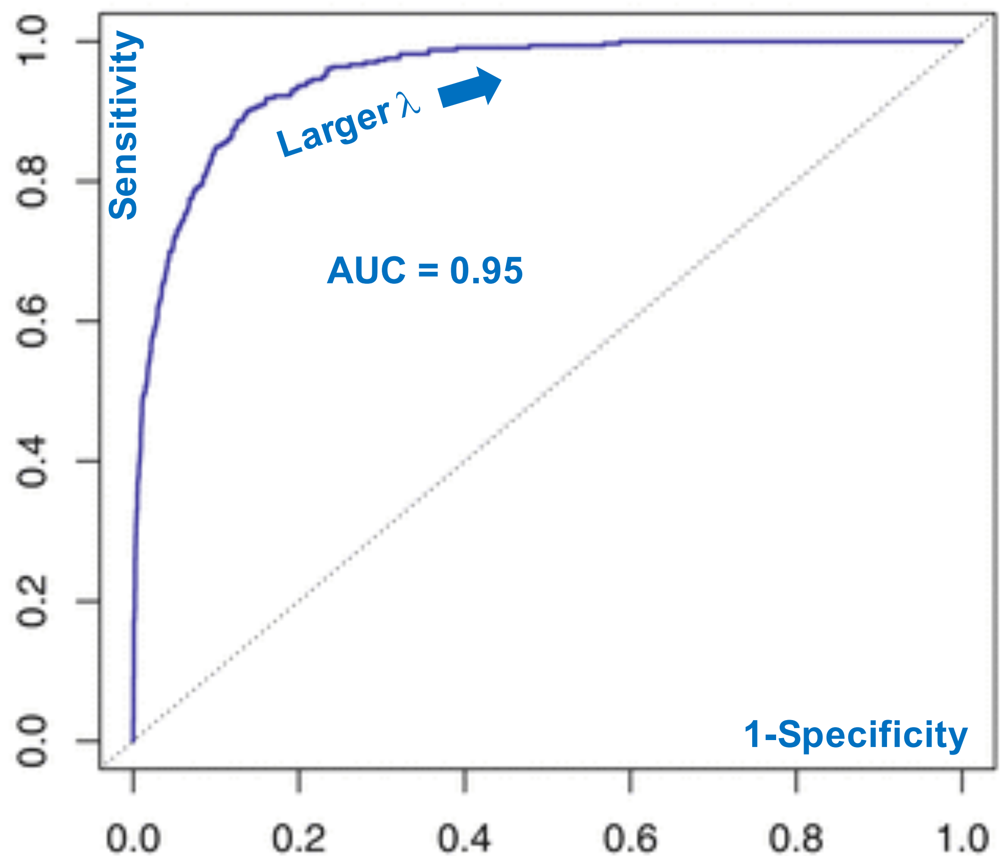
- **Error Rate** ↑; **Sensitivity** ↑; **Specificity** ↓
- Importantly, the **prediction of defaults** has improved from **24.3%** to **58.6%** (better than flipping a coin)
- Again, **tuning** the model to our analysis goals is **key!!** If your business goal is to predict defaults, $\lambda = 0.2$ is **better** than 0.5

How to Select λ ?

- In any **classification** model based on **probability** (e.g., logistic regression, LDA), you can vary the **classification probability** λ and get different results
- So, what is the **best value** of λ?
- The best approach is to try **severa**l values of λ from $\mathbf{0}$ to $\mathbf{1}$ and comparing the resulting **Sensitivity** and **Specificity** values
- Naturally, a model that provides **large values** of both, **Sensitivity** and **Specificity** at various values of λ is the **best model** because it provides more accurate classification
- The specific λ to select within that model to select depends on the **analysis goals**
- As $\lambda \uparrow$ increases **Error Rate** $\downarrow$;(overall); **Sensitivity** $\downarrow$; Specificity $\uparrow$ ith λ. This is a **tradeoff** → $\downarrow$ overall Error → $\downarrow$ false negatives; $\uparrow$ false positives.
- **ROC** curve (“Receiver Operating Characteristics”) helps **visualize** this tradeoff

ROC Curve

- The **ROC** is a plot constructed by trying **several** values of λ from **0** to **1** in the model and plotting the resulting **Sensitivity** and **1-Specificity**.
- The larger the “area under the curve” (**AUC** close to **1**) the **better** the model – i.e., the curve **“hugs”** the top-left corner.
- The **dotted** 45^0 line represents a model that performs no better than chance with **AUC = 0.5**



Kogod School
of
Business