DSO530 Statistical Learning Methods

Lecture 3c: Classification III

Dr. Xin Tong
Department of Data Sciences and Operations
Marshall School of Business
University of Southern California

Thresholds

- The default threshold for $P(Y = 1|X = x)$ is usually 0.5
- If we were to increase the threshold to 0.8, how will type I error and type II error change?
- How about decreasing the threshold to 0? What will happen?
- A takeaway message: for binary classification, you can always trade-off one error at the expense of the other.
- Q: Is the problem of choosing a proper threshold **unique** to logistic regression?

Thresholds

- The default threshold for $P(Y = 1|X = x)$ is usually 0.5
- If we were to increase the threshold to 0.8, how will type I error and type II error change?
- How about decreasing the threshold to 0? What will happen?
- A takeaway message: for binary classification, you can always trade-off one error at the expense of the other.
- Q: Is the problem of choosing a proper threshold **unique** to logistic regression?

A principled way to balance type I and type II errors

- How to achieve type I error / type II error control on the population level? No easy
- (Optional) For a non-statistician's introduction of **the Neyman-Pearson classification paradigm**: https://fxdiebold.blogspot.com/2019/03/neyman-pearson-classification.html
- (Optional) An introduction of the **Neyman-Pearson (NP) umbrella algorithm** which adapts the main-stream classification methods to the NP paradigm: https://medium.com/@ruhandon/summary-for-neyman-pearson-classification-algorithms-a0c9595632a9
- (Not recommended for reading unless you have a master degree in Statistics) Reference: https://advances.sciencemag.org/content/4/2/eaao1659/tab-pdf

Linear discriminant analysis

- Linear discriminant analysis (LDA) is a model that appears often in the statistics literature for its nice mathematical properties to analyze
- We skip its details here because it usually has a similar empirical performance to logistic regression, and in practice, it has far less popular compared to logistic regression
- However, it is worth to know that from the modeling perspective, LDA and logistic regression are two different approaches.
- What is an LDA model:

$$X|(Y=0) \sim \mathcal{N}(\mu_0, \Sigma) \quad \text{v.s.} \quad X|(Y=1) \sim \mathcal{N}(\mu_1, \Sigma)$$

- Key: two class normal; different means, but same covariance matrix
- How about model fitting? The means and covariance are unknown parameters. We use the so-called **plug-in approach** (not required).
- Recall the logistic regression model. Can you name one difference between the LDA model and the logistic regression model?