DS 102 Discussion 1

Wednesday, September 2, 2020

1. **ROC Curves.** In lecture we defined and discussed ROC curves, or "receiver operating characteristic" curves. ROC curves plot the true positive rate (TPR) and false positive rates (FPR) for a binary classifier at different decision thresholds. Recall that the TPR and FPR are defined as:

$$\text{TPR} = \frac{\#\text{ true positives}}{\#\text{ positives}}, \quad \text{FPR} = \frac{\#\text{ false positives}}{\#\text{ negatives}},$$

where "true positives" are examples where the model made a positive decision and the label was positive, and "positives" are examples where the label was positive.

In this exercise, we will consider the ROC curve on an example dataset. Let Y be the label, X_1, X_2 be features, and consider the model function $f(X_1, X_2) = 3X_1 + 2X_2 + 1$.

Table 1: Example dataset

Y	$f(X_1, X_2)$	X_1	X_2
0	-1	-1	0.5
1	-0.5	-1	0.75
0	0	-1	1
1	1	0.2	-0.3
1	0.25	-0.25	0
0	0.25	-0.05	-0.3

(a) Plot the ROC curve for the model $f(X_1, X_2)$ with respect to the label Y.

(b) Suppose that we can choose two decision thresholds α_1 and α_2, and for each data example, we flip a coin to decide which decision threshold to use for that example. Choose α_1 and α_2, and probabilities for using α_1 and α_2, such that in expectation, the true positive rate is $\frac{1}{3}$ and the false positive rate is $\frac{1}{3}$.

(c) Is it possible to choose two decision thresholds α_1, α_2 and probabilities of using each decision threshold such that the expected true positive rate is $\frac{1}{3}$, and the expected false positive rate is $\frac{2}{3}$?

2. **Hypothesis Testing.** As discussed in lecture, one can imagine different metrics for quantifying how "good" a decision is. For example, we would like our decisions to have both high true positive rate and low false positive rate. Our goal as statisticians is to develop reasonable strategies for doing well on both metrics. In other words, how should we pick a point on the ROC curve? Once we pick a point, how do we achieve it?

   The Neyman-Pearson Lemma offers one solution. To be concrete, we focus on the case of hypothesis testing. We call the probability of a false positive under null hypothesis H_0 the *significance level* α of a test, and we call the probability of a true positive under the alternative hypothesis H_1 the *power* of a test.

   The Neyman-Pearson formulation prescribes the following point on the ROC curve: fix a significance level you are willing to tolerate, then pick the point that maximizes power. The Neyman-Pearson Lemma prescribes how to achieve this point:

   Lemma (Neyman & Pearson, 1933) *Suppose $\theta_1 < \theta_0$. For any significance level $\alpha \in [0,1]$, the following* likelihood-ratio test *maximizes power among all tests with level at most α:*

$$\delta(x) = \begin{cases} \textit{Reject Null} & : \ \frac{f_{\theta_0}(x)}{f_{\theta_1}(x)} \leq \eta \\ \textit{Accept Null} & : \ \frac{f_{\theta_0}(x)}{f_{\theta_1}(x)} > \eta \end{cases}$$

   where $f_{\theta_0}, f_{\theta_1}$ are the likelihoods under the null and alternative distributions, respectively, and η is the real value such that $Pr(\delta(X) = 1 \mid H_0) = \alpha$.

Example. Suppose that you have a sample from a distribution with probability density function $f_\theta(x) = \theta x^{\theta-1}$ where $0 < x < 1$. You would like to design a test to discern between the null hypothesis that $\theta = 4$, and the alternative hypothesis that $\theta = 3$.

(a) Derive the most powerful test for this problem such that the significance level is less than α.

(b) What is the power of the test, $Pr(\delta(X) = 1 \mid H_1)$?