IEOR 190D/290D Data-X: Data, System, and Signals

Handout (Lecture 6)

February 22, 2017

1 Loss Functions

Loss Function: $L(\theta, \delta(X))$

Expected Loss: $R(\theta, \delta) = \mathbb{E}(L(\theta, \delta(X)))$

Question:

Why is the Loss Function hard to evaluate?

Ans:

1.1 Classical Loss Functions

0-1 Loss

$$L(\theta) = \begin{cases} 0 & \text{if correct} \\ 1 & \text{otherwise} \end{cases}$$

$$R(\theta, \delta(X)) = \begin{cases} R(0, d=1) = P_0(d=1) & \text{if } \theta = 0 \\ R(1, d=0) = P_1(d=0) & \text{if } \theta = 1 \end{cases}$$

Squared-Error Loss

$L(\theta, \delta(X)) = (\theta - \delta(X))^2$

$R(\theta, \delta(X)) = \mathbb{E}_\theta(g(\theta) - \delta(X))^2$

2 An Example: Coin Tosses

Consider the following example:

For a coin toss, the chance of heads is very close to $\frac{1}{2}$. Suppose instead we stand a coin on its edge, balancing it with a finger on top, and spin it by flicking it with a different finger. If this is done 100 times, with the trials independent and a common chance θ of heads on each spin, then the total number of heads X should have a Binomial distribution.

In particular, $X \sim Bin(100, \theta)$

Viewing X as our data and taking $P_\theta = \text{Binomial}(100, \theta)$, and $\theta \in [0, 1] = \Omega$.

In this example, a natural estimator of θ is $\delta(X) = X/100$.

The question we want to ask here is: How well does the estimator do?

i.e. How does $R(\theta, \delta)$ behave?

Consider the squared error loss,

The risk function for δ is: $R(\theta, \delta) = \mathbb{E}_\theta(\theta - X/100)^2 = \frac{\theta(1-\theta)}{100}$

Question: How well does the sample mean as an estimator (decision $\delta(X)$ do?

Let's consider some alternative decision/estimators:

1. $\delta_0(x) = x/100$; $R(\theta, \delta_0) = \theta(1-\theta)/100$
2. $\delta_1(x) = (x+3)/100$; $R(\theta, \delta_1) = (9 + 100\theta(1-\theta))/100^2$
3. $\delta_2(x) = (x+3)/106$; $R(\theta, \delta_2) = (9 - 8\theta)(1 + 8\theta)/106^2$

3 Applications in Real Life

1. Spam Filter
2. Medical Diagnosis
3. Signal Detection (Military)
4. Finance: Trading/Forecasting
5. Online Advertisements