Introduction, Notation, and Overview

-Introduction to Data Science-

ISL: Chapter 2

Lecturer: Darren Homrighausen, PhD

Risk, Bayes, bias, variance, and approximation

Loss functions and risk

If we want a $\hat{f}(X)$ which is a good prediction, what does good mean?

Define a loss function which

- Inputs both
  - $\underbrace{\hat{f}(X)}_{\text{Our prediction}}$
  - $\underbrace{Y}_{\text{Unknown, true value}}$
- Outputs a number $\ell(\hat{f}(X), Y)$ between 0 and ∞...

...such that smaller $\ell(\hat{f}(X), Y)$ indicate better performance

(There is an intimate connection between loss and likelihoods, hence same notation)

Risky (and lossy) business

Any distance function could serve for the loss function ℓ

As both $\hat{f}(X)$ and Y are random, the loss function is random

Hence, we define the risk to be the expectation of the loss

$$R(f) = \mathbb{E}\ell(f(X), Y)$$

(Hence, the risk is not random)

Definition: A good procedure f is one that has a small risk $R(f)$

Example: Wireless network

Example

We run a large wireless network with SMS
(SMS: Short Message Service)

A user is entering text to send to another user

Using the semantics of the text, we have identified that a phrase has ended and hence either a "!" or "." should follow

The phrase is Thank you

Females and males use "!" at very different rates. Luckily, we know the sender's gender (Male) as well

Hence, our $X = [\text{Thank you}, \text{Male}]$

Example

Two possible procedures: Search over the training data $\mathcal{D}$ and find:

$$\hat{\pi}_1(X) = \frac{\#\text{ times ``!'' follows ``you'' and sender is male}}{\#\text{ times ends in ``you'' and sender is male}}$$

$$\rightarrow \hat{f}_1(X) = \begin{cases} \text{``!''} & \text{if } \hat{\pi}_1(X) > \frac{1}{2} \\ \text{``.''} & \text{otherwise} \end{cases}$$

Or

$$\hat{\pi}_2(X) = \frac{\#\text{ times ``!'' follows ``thank you'' and sender is male}}{\#\text{ times ends in ``thank you'' and sender is male}}$$

$$\rightarrow \hat{f}_2(X) = \begin{cases} \text{``!''} & \text{if } \hat{\pi}_2(X) > \frac{1}{2} \\ \text{``.''} & \text{otherwise} \end{cases}$$

Which procedure is better?

Example

Using the loss/risk idea, we should define a loss function

A sensible loss for this problem would mimic the question:

$$\text{Did I get the punctuation right?}$$

We can quantify this via the 0 - 1 (e.g. zero-one) loss function

$$\ell(\hat{f}(X), Y) = \mathbf{1}(\hat{f}(X) \neq Y) = \begin{cases} 0 & \text{if } \hat{f}(X) = Y \\ 1 & \text{if } \hat{f}(X) \neq Y \end{cases}$$

where $\mathbf{1}(A)$ is an indicator function

$$\mathbf{1}(A) = \begin{cases} 1 & \text{if statement } A \text{ is true} \\ 0 & \text{if statement } A \text{ is false} \end{cases}$$

Example

Suppose we have data

	"!"	"."
"$*$ you"	3712	20463
"thank you"	2003	1012

Then $\hat{\pi}_1 \approx 0.154 \rightarrow \hat{f}_1(X) =$ "." and $\hat{\pi}_2 \approx 0.664 \rightarrow \hat{f}_2(X) =$ "!"

If $Y =$ "!", then the loss value for each procedure is:

$$\ell(\hat{f}_1(X), Y) = 1$$

$$\ell(\hat{f}_2(X), Y) = 0$$

If $Y =$ ".', then the loss value for each procedure is:

$$\ell(\hat{f}_1(X), Y) = 0$$

$$\ell(\hat{f}_2(X), Y) = 1$$

Which procedure do we prefer? The one with lower risk!

Example

Suppose the true data generating process is that the probability that a male user uses an "!" after "Thank you" is 0.57.

What is the risk of each procedure?

$$R(\hat{f}_1) = \mathbb{E}\ell(\hat{f}_1(X), Y) = ?$$
$$R(\hat{f}_2) = \mathbb{E}\ell(\hat{f}_2(X), Y) = ?$$

(Refer to the homework)