CS 4980/6980: Introduction to Data Science © Spring 2018

Lecture 9: Introduction to Information Theory

INSTRUCTOR: DANIEL L. PIMENTEL-ALARCÓN Scribed by: Xiang Li

This is preliminary work and has not been reviewed by instructor. If you have comments about typos, errors, notation inconsistencies, etc., please email the scribers.

9.1 Anouncement

The midterm exam is in next monday(10:00AM-11:45AM). The content of the exam will cover:

1. Linear Regression.
2. Logistic Regression.
3. K-means Custering.
4. Nearest Neighbers.
5. Linear Algebra.
6. Probability.

No cheatsheet in the exam. Belows are some hints:

1. Be able to multiply vectors, matrices.
2. Be able to invert a matrice.
3. Be able to compute norms.

The lecture of this Wednesday will review and answer questions.

9.2 Information Theory

Information theory is about:

1. Quantify/study how much informaiton is contained in data.
2. How to store data efficiently.
3. How to communicate data/information reliable.

Example 9.1 (Discussion)**.** Imaging that there is a horse race with 8 horses. The winning probabiliy of these horses are 1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64. Please find 8 codes to represent these 8 horses so that when we store the code in database, the cost of memory will be minimum.

horse number	1	2	3	4	5	6	7	8	E(L)
P(winning)	1/2	1/4	1/8	1/16	1/64	1/64	1/64	1/64	E(L)
example code	000	001	010	011	100	101	110	111	3 bits/race
XX code	1	01	001	0001	000000	000001	000010	000011	2 bits/race
Jason's code	1	01	001	0001	00001	000001	0000001	00000001	2.031 bits/race

Here is how we calculate the expectation of length:

$$
\begin{aligned}
E[length\ of\ naive\ code] &= \sum_{h=1}^{8} p(h) * length(code\ of\ house\ h) \\
&= \frac{1}{2}(3) + \frac{1}{4}(3) + \frac{1}{8}(3) + \frac{1}{16}(3) + \frac{1}{64}(3) + \frac{1}{64}(3) + \frac{1}{64}(3) + \frac{1}{64}(3) \\
&= 3 \\
E[length\ of\ XX\ code] &= \sum_{h=1}^{8} p(h) * length(code\ of\ house\ h) \\
&= \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{16}(4) + \frac{1}{64}(6) + \frac{1}{64}(6) + \frac{1}{64}(6) + \frac{1}{64}(6) \\
&= 2
\end{aligned}
$$

Question: Do some codes exist that their efficiencies are less than 2?

Answer: No. We can prove this from Entropy.