Lesson Plan 3/13

Monday, March 12, 2018 3:29 PM

Admin
- Project 1 due tomorrow 3/14 24:00
- Change in grading weights

HDSC
- CBCB summer internship program: https://www.cbcb.umd.edu/summer-internships

Project 1
- Outstanding questions and comments
- Update to two table example (how to turn similarity matrix into data frame)

HW3
- Go over description
- Work on #1

Intro stats
- Random variables
- Discrete probability distributions
- Expectation

Estimation
- LLN
- CLT
- Normal distribution
- Continuous probability distributions
- CLT finalized

Homework 3, Q2

(a) Derive $\bar{Z}$

(b) Derive S_Z

$$\bar{z} = \frac{1}{n} \sum_i z_i$$
$$= \frac{1}{n} \sum_i \frac{x_i}{s_x}$$
$$= \frac{1}{s_x} \cdot \left[\frac{1}{n} \sum_i x_i \right]$$
$$= \frac{\bar{x}}{s_x}$$

$$s_z^2 = \frac{1}{n} \sum_i (z_i - \bar{z})^2$$
$$= \frac{1}{n} \sum_i \left(\frac{x_i}{s_x} - \frac{\bar{x}}{s_x} \right)^2$$
$$= \frac{1}{s_x^2} * \left[\underbrace{\frac{1}{n} \sum_i (x_i - \bar{x})^2}_{s_x^2} \right]$$
$$= \frac{1}{s_x^2} * s_x^2$$
$$= 1$$

→ Population vs. sample

→ Notation & properties

→ Distributions of discrete variables

→ continuos

→ Central limit theorem

→ Expectation

Data: Tweets (about a specific topic or hashtag)

→ Bot or human

Entities: tweets
Attribute: hot or not

$x_i \in \{0, 1\}$ ← observed

$X_i \in \{0, 1\}$ ← random variable

→ Probability distr. of X_i

$P : D \to [0, 1]$

D : values that X_i can take

(a) density
$p(X_i = x_i) > 0$ $\qquad x_i \in D$

(b) $\sum_{D} p(X_i = x_i) = 1$

$\Rightarrow$ Oracle

$$p(X_i = 1) = .7$$

$\Rightarrow$ Expectation

$$E X_i = \sum_{x_i \in D} x_i \, p(X_i = x_i)$$

$$E X_i = \underline{0 * p(X_i = 0)} + \underline{1 * p(X_i = 1)}$$

$= .7$

$\Rightarrow$ Estimation

1. Compute sample mean
2. set equal to expected value
3. solve!

$n = 100$

$X_1 \in \{0,1\}, \ldots \quad X_n \in \{0,1\}$

1. $\frac{1}{n} \sum_i X_i$

(2) $E\left[\frac{1}{n}\sum_i X_i\right] = \frac{1}{n}\sum_i E\,X_i = \frac{1}{n}\sum_i p = p$

(3) $\hat{p} = \bar{x}$

Estimate