Lesson Plan 4/3

Tuesday, April 3, 2018 12:11 PM

Admin
- Project 2 due on Friday
- HW4 released (due on Tuesday)
- Final project posted
- Grades for Project 1 and HW3 almost done (looking good!)

HDSC:
- Mobility charts from The Upshot: https://www.nytimes.com/interactive/2018/03/27/upshot/make-your-own-mobility-animation.html?rref=collection%2Fsectioncollection%2Fupshot&action=click&contentCollection=upshot&region=rank&module=package&version=highlights&contentPlacement=7&pgtype=sectionfront

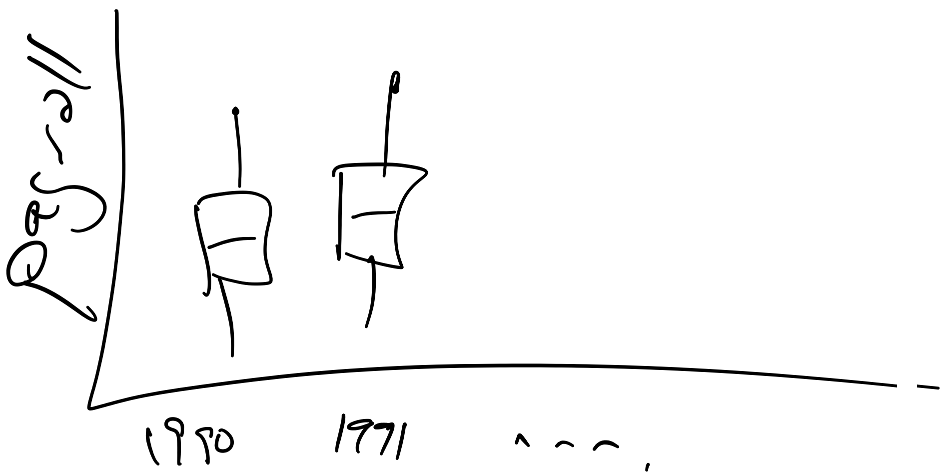
Project 2 Questions?
Final Project Description
HW4
- Discussion
- Example derivation

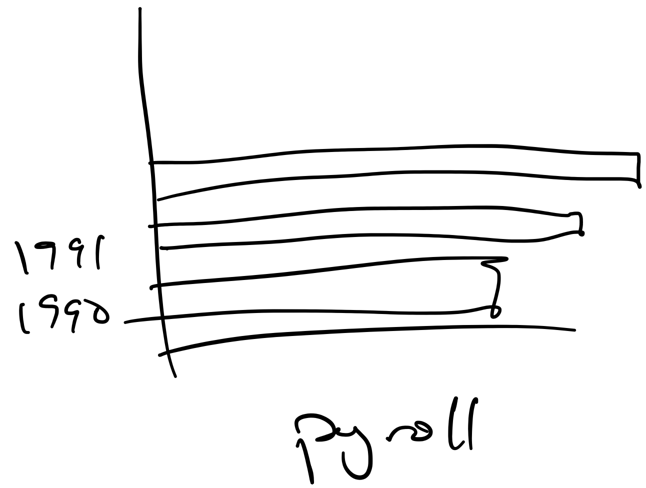
Bayes Rule
Data Analysis with Geometry
- Preliminaries
- Geometry and distances
- KNN classifier
- Some vector algebra

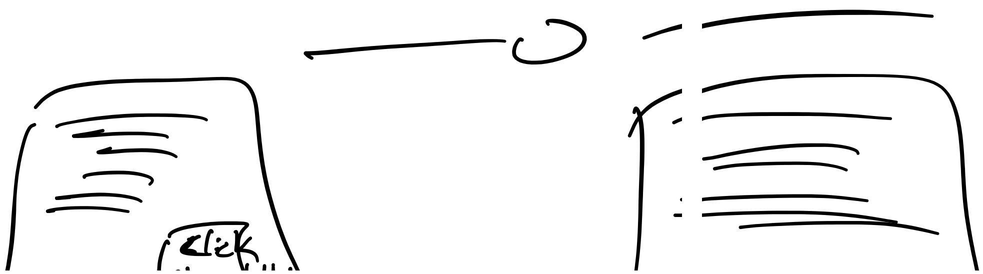
Proj 2 Problem 2

year	team ID	payroll	


payroll (M$)


payroll
1990
1991


1991
1990
payroll


click

A B

Clickrate p_A Click rate p_B

Part 1) known $p_A = .5$

Part 2) known $p_A = 0.75$

Part 3) – – – –

part 4) Both p_A and p_B are estimated

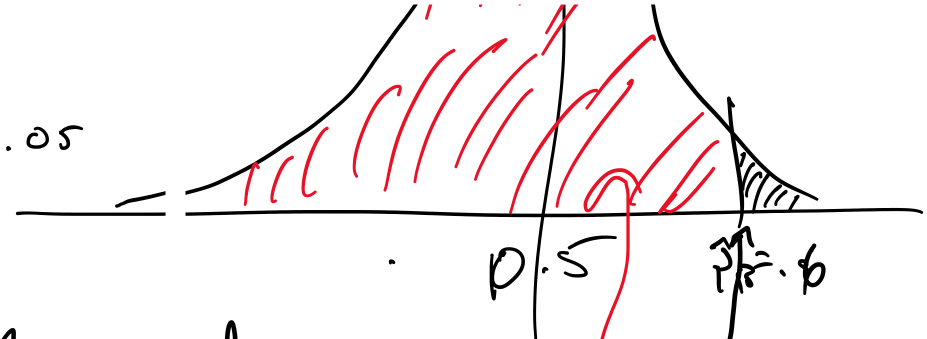
Hypothesis: $p_B \le p_A$

Estimate $\hat{p}_B = .6$

Estimate $\hat{p}_B$

if $p(\bar{X} > \hat{p}_B) \le .05$

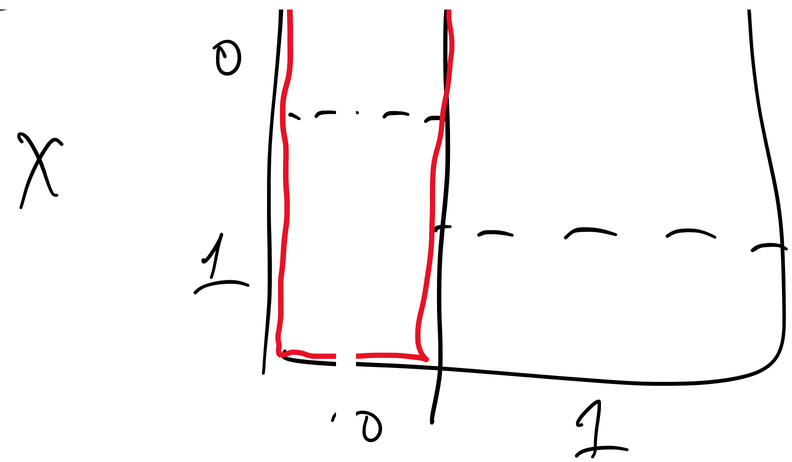
then Reject



Approximate as $N(0.5, \sqrt{var(\bar{X})})$

```
pnorm(.6, mean=0.5, sd= ....)
qnorm(.95, mean=0.5, sd= ....)
```

Joint & Conditional Probability



$p(x,y)$

(a) $p(X=x, Y=y) \geq 0 \quad \forall (x,y) \in D_x \times D_Y$

(b) $\sum_{(x,y)\in D_x \times D_Y} p(X=x, Y=y) = 1$

Conditional

$$P(X=x \mid Y=0) = \frac{p(X=x, Y=0)}{p(Y=0)}$$

Independence

$$P(X=x \mid Y=y) = P(X=x) \text{ for all } y \in D_Y$$

Easier to think about $P(Y=y \mid X=x)$ than the opposite $P(X=x \mid Y=y)$

Sentiment Analysis

X: {positive vs. negative}

Y: word frequency

Bayes Rule:

A way of exchanging conditional probabilities

$$p(X=x \mid Y=y) = \frac{p(Y=y \mid X=x)\, p(X=x)}{p(Y=y)}$$

$$p(X=x \mid Y=y) = \frac{p(X=x, Y=y)}{p(Y=y)}$$

$$= \frac{p(Y=y \mid X=x)\, p(X=x)}{p(Y=y)}$$

$$p(X=x \mid Y=y) = \underbrace{p(Y=y \mid X=x)}_{\text{data distribution}} \; \underbrace{p(X=x)}_{\text{prior distribution}}$$

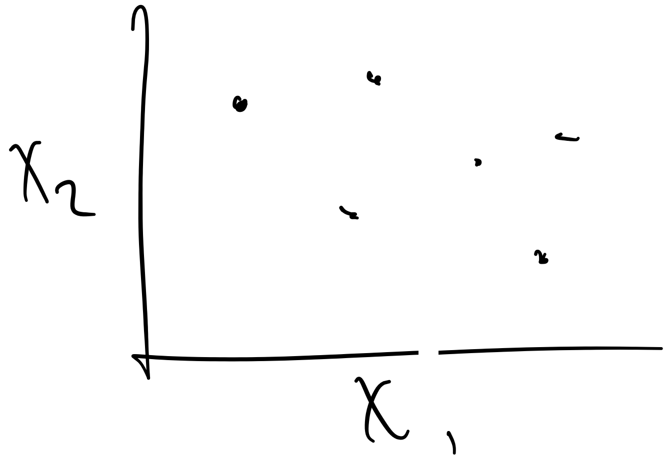
$$\underbrace{p(X=x \mid Y=y)}_{\text{posterior distribution}} \qquad \frac{\quad}{p(Y=y)} \Leftarrow \text{normalizing constant}$$

start thinking about data as geometric objects

$$p(\text{default} = \text{yes} \mid \text{student}, \text{balance}, \text{income})$$

outcome: y

predictors: $x_1, x_2, \ldots, x_n$



Predictors