

## Homework 2

CS 436/: Introduction to Machine Learning

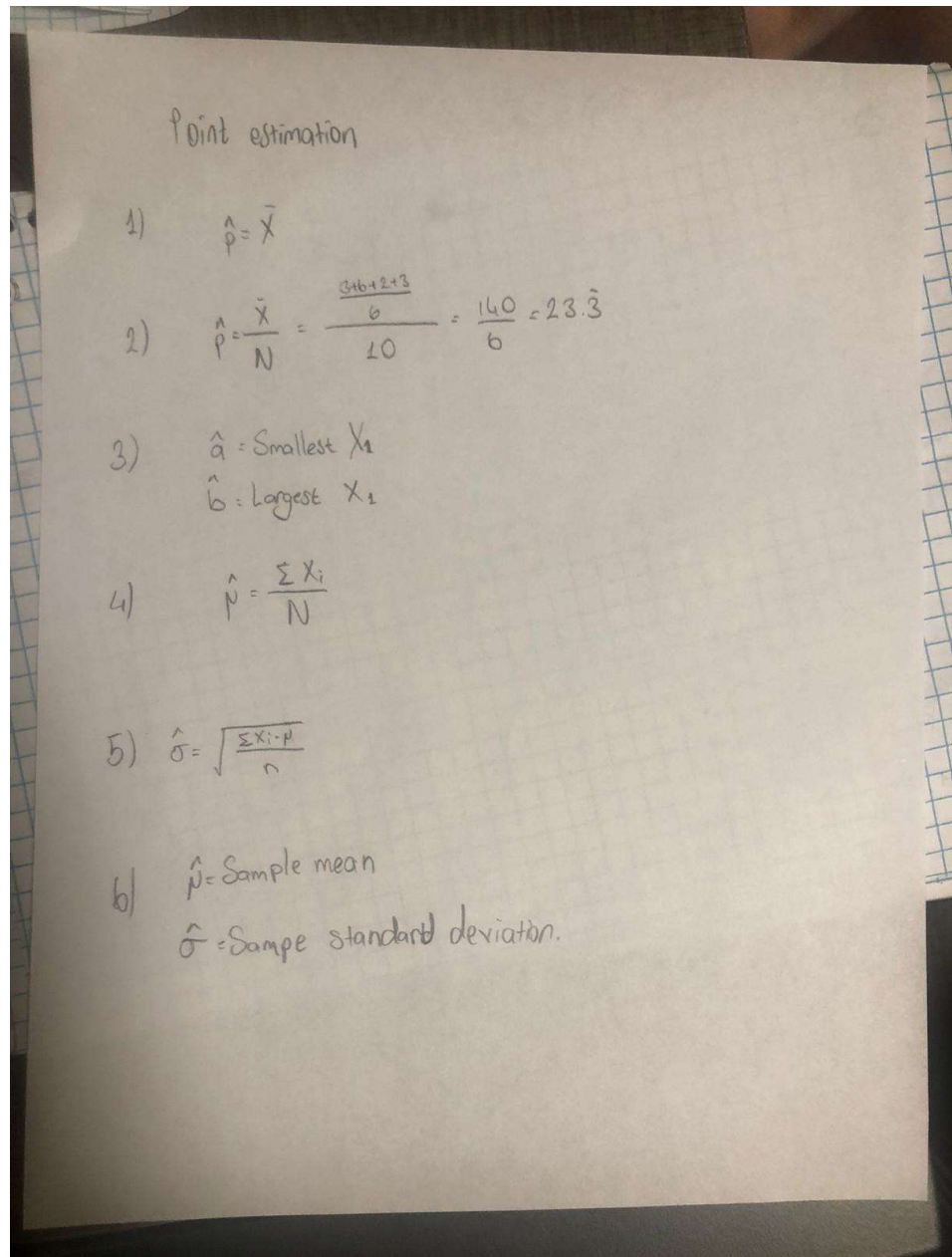
Egemen Inal

### 1) Naive Bayes for Text Classification

Accuracy on the test file with Stop words is %96.44351464435147

Accuracy on the test file without Stop words is % 94.56066945606695

### 2) Point Estimation



4 Coin

60H, 40T

Thumbtack

70H, 30T

Betas

1,1

40,60

80,20

100,00

1000,1000

100,000,100,000

MLE - Coin - Heads

60

100

MLE

Thumbtack

70

100

Heads

40

100

Heads

30

100

MAP - COIN -

$$= \frac{60+1-1}{60+40+1+1-2} = \frac{60}{100}$$

$$= \frac{160-1}{800-2} = \frac{159}{798}$$

$$= \frac{60+40-1}{60+40+60+40-2} = \frac{99}{198}$$

$$= \frac{1060-1}{2100-2} = \frac{1059}{2098}$$

$$= \frac{60+80-1}{200-2} = \frac{139}{198}$$

$$= \frac{109,060-1}{200,100-2} = \frac{109,059}{199,998}$$

MAP - Thumbtack

$$= \frac{70}{100}$$

$$= \frac{109}{198}$$

$$= \frac{39}{198}$$

$$= \frac{169}{298}$$

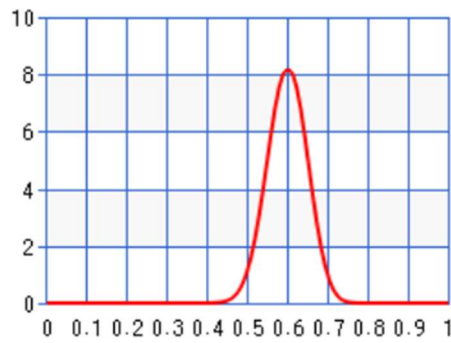
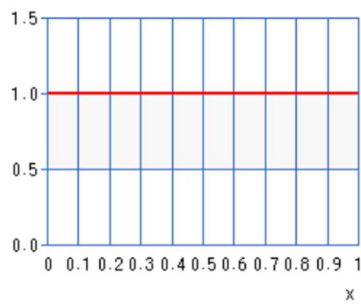
$$= \frac{1069}{2098}$$

$$= \frac{100,069}{200,098}$$

2)

Prior

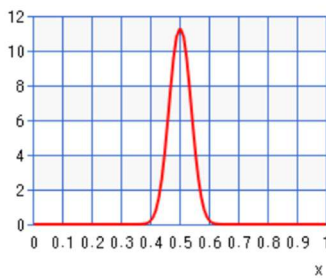
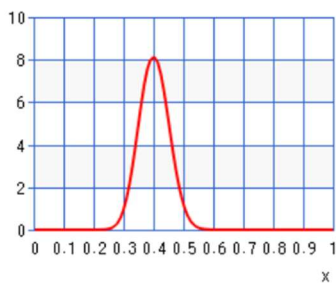
Likelihood(coin)



Prior is not useful and close to 0.5.

Prior

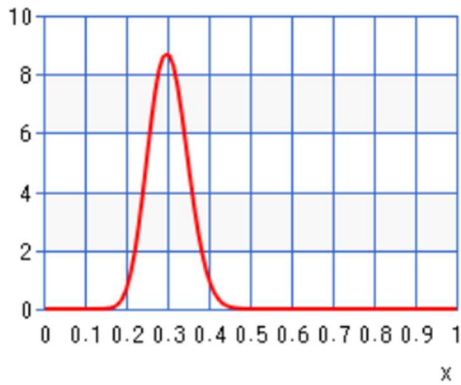
Likelihood(coin)



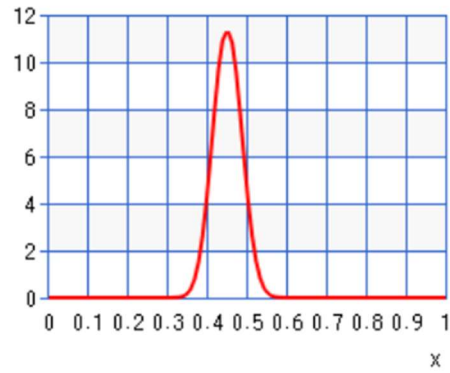
Both are at 0.5 they both contribute equally.



Prior

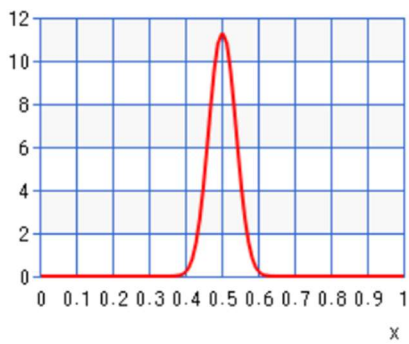


Likelihood(coin)

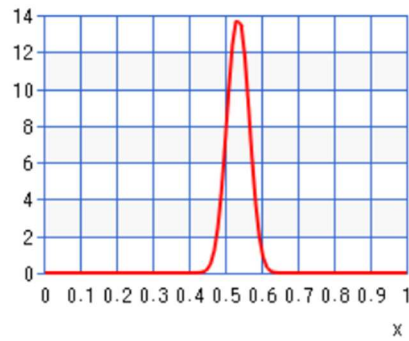


Both contribute equally and likelihood is closer to 0.5.

Prior

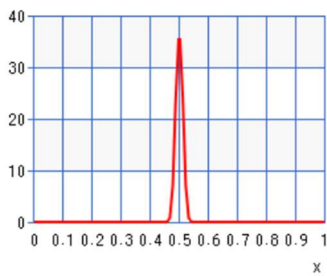


Likelihood(coin)

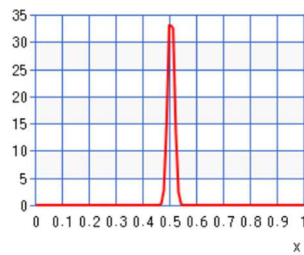


Prior contributes more, gets closer to 0.5 and gets narrower.

Prior

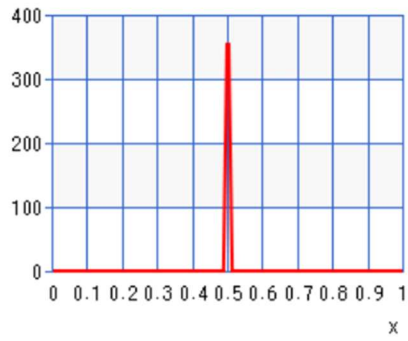


Likelihood(coin)

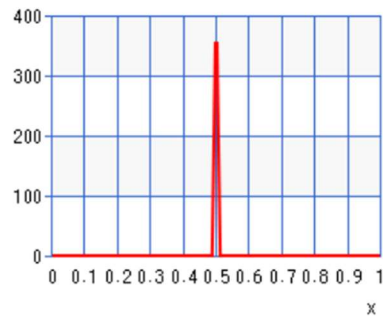


Prior contributes more, gets closer to 0.5 and gets narrower.

Prior

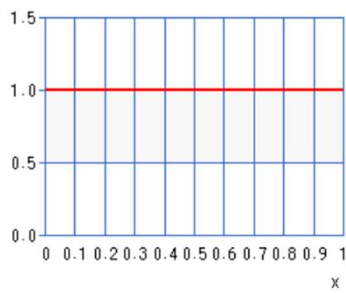


Likelihood(coin)

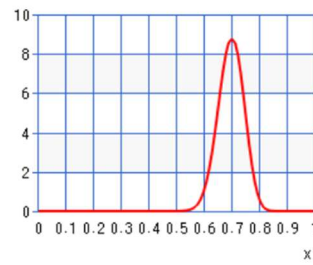


Prior contributes more, gets closer to 0.5 and gets narrower.

Prior

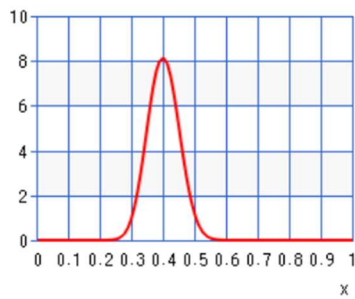


Likelihood(thumbtack)

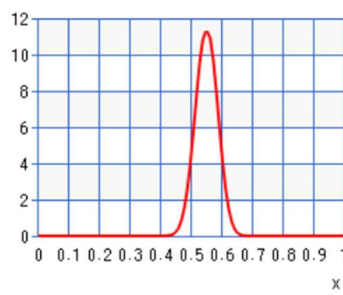


Prior is not useful and close to 0.5.

Prior

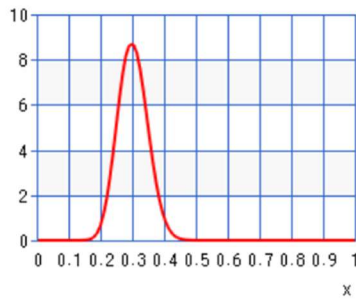


Likelihood(thumbtack)

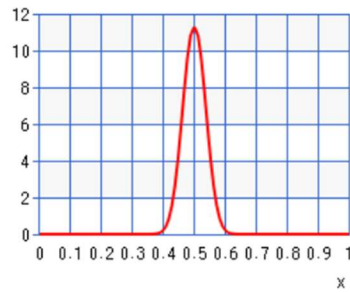


Both are at 0.5 they both contribute equally.

Prior

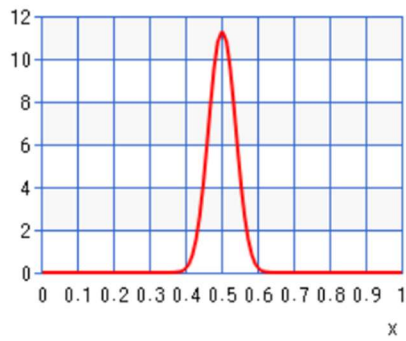


Likelihood(thumbtack)

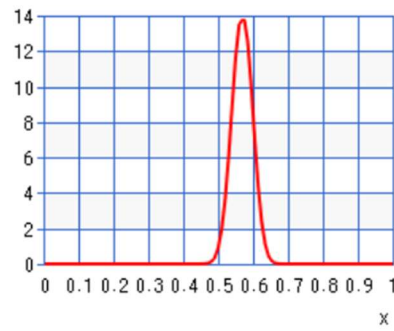


Both contribute equally and likelihood is closer to 0.5.

Prior

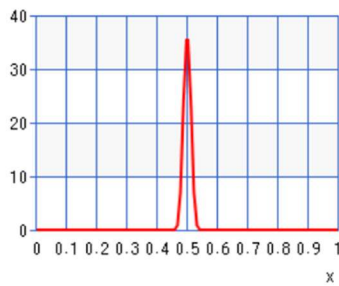


Likelihood(thumbtack)

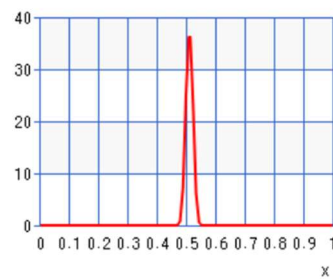


Both contribute equally and likelihood is closer to 0.5.

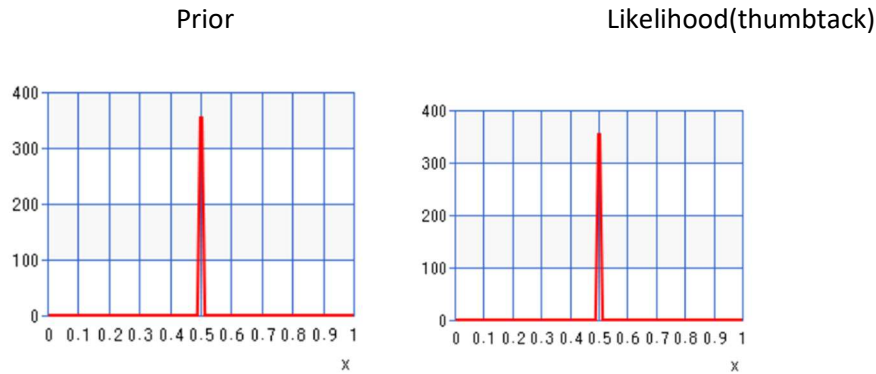
Prior



Likelihood(thumbtack)



Both contribute equally and likelihood is closer to 0.5.



Both contribute equally and likelihood is closer to 0.5.

3) True, Data Dominates

4) True, Prior Dominates

I have done this assignment completely on my own. I have not copied it, nor have I given my solution to anyone else. I understand that if I am involved in plagiarism or cheating I will have to sign an official form that I have cheated and that this form will be stored in my official university record. I also understand that I will receive a grade of 0 for the involved assignment for my first offense and that I will receive a grade of "F" for the course for any additional offense.

Egemen Inal