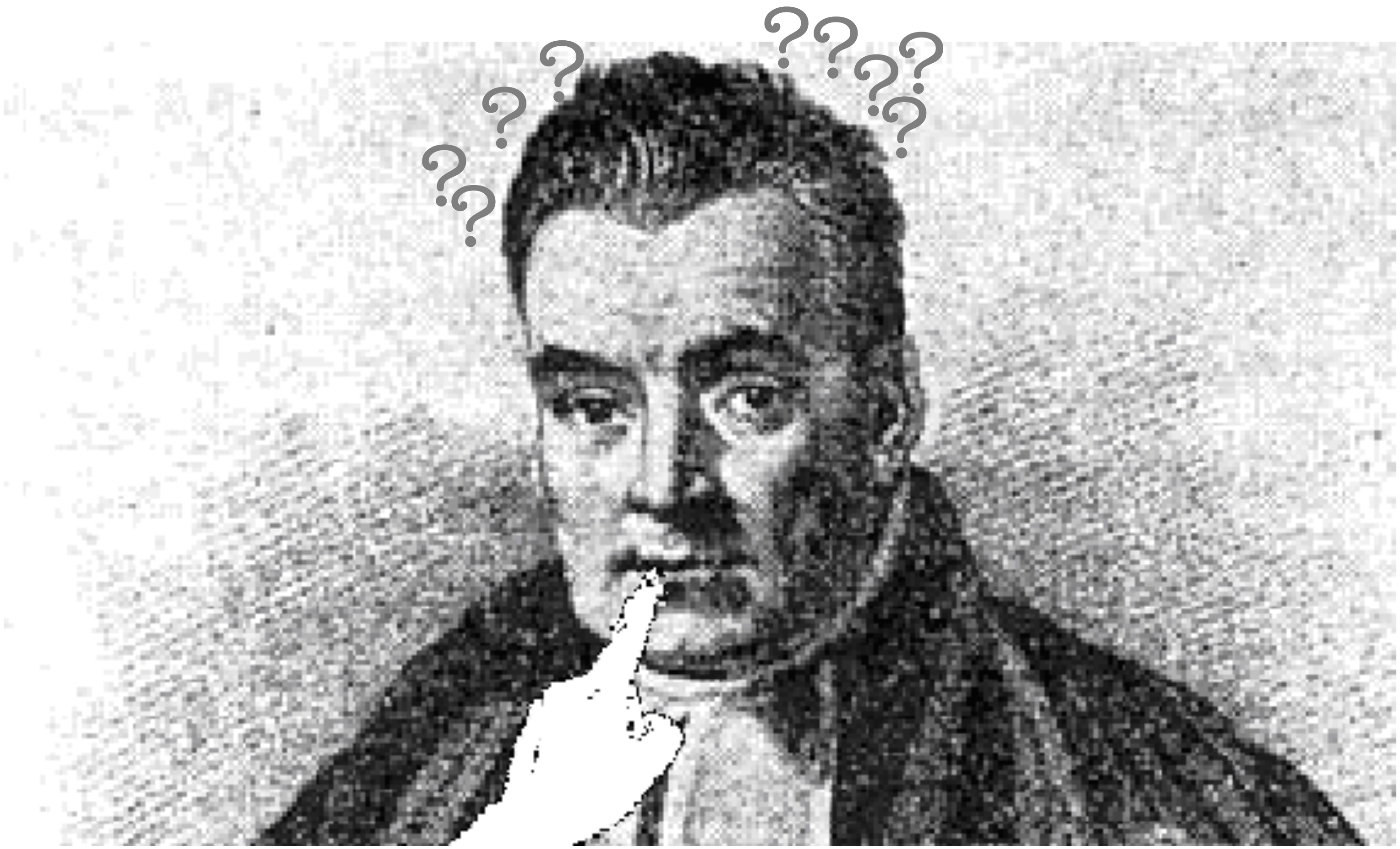
Naive Bayes

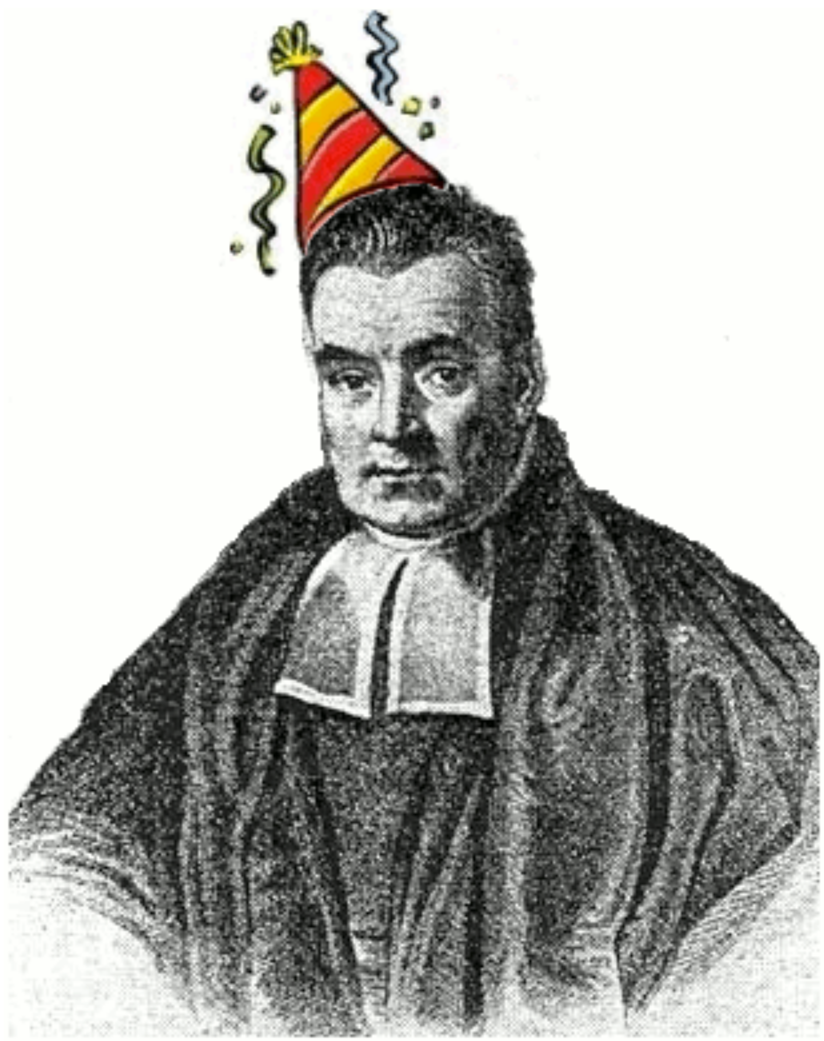
Naïve



Updating the state of knowledge

step by step

with new information



What is classification?

Deciding among hypotheses (labels),
using information we have (features)
for each example

3 Features: Votes on 3 Bills

2 Labels: Democrat / Republican

Prediction:
I know your votes, I'm trying to guess your party

3 Features: Votes on 3 Bills

2 Labels: Democrat / Republican

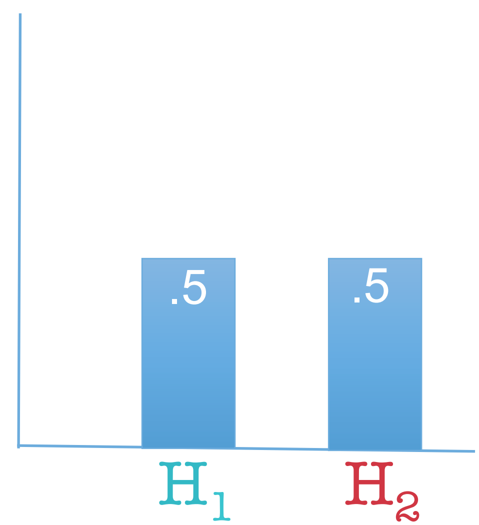
Prediction:

I know your votes, I'm trying to guess your party

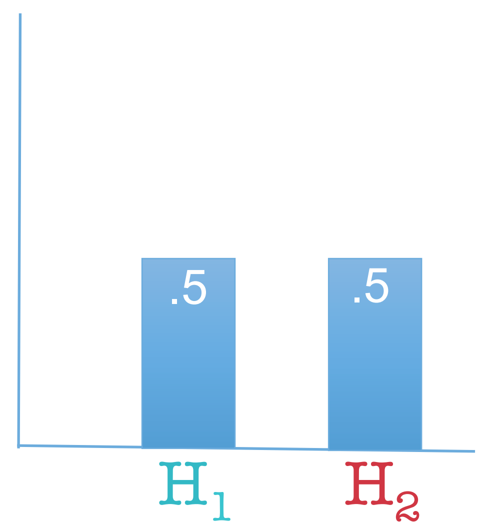
2 Labels

H_1: Democrat

H_2: Republican



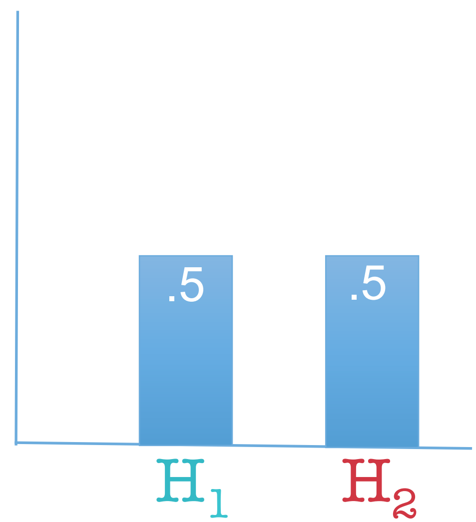
Prior: Initial belief



Prior: Initial belief

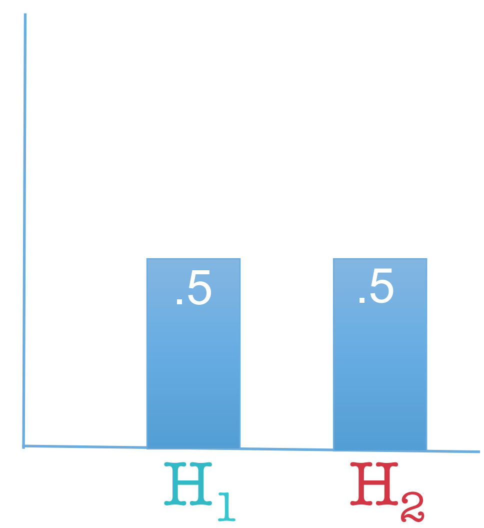
50 - 50 ?   P(Democrat) = 0.5 ?

(Uninformative prior)



Prior: Initial belief

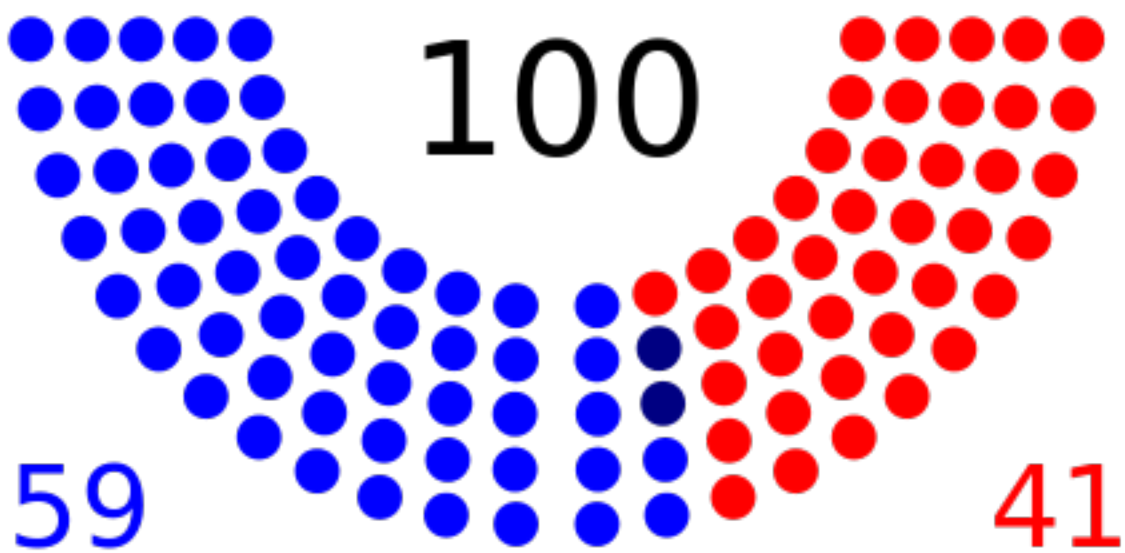
What's my best guess without any vote info?

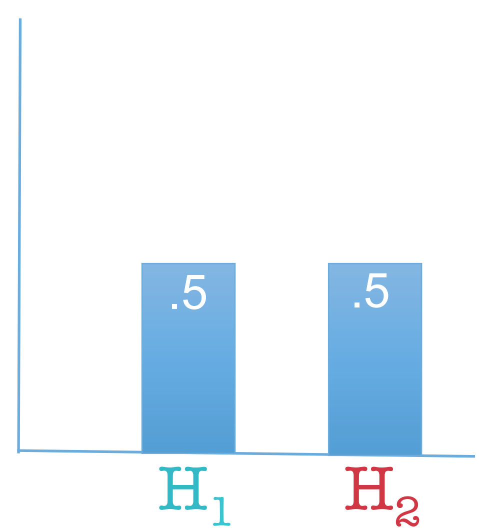


Prior: Initial belief

What's my best guess without any vote info?

I'd guess democrat since there are more of them.



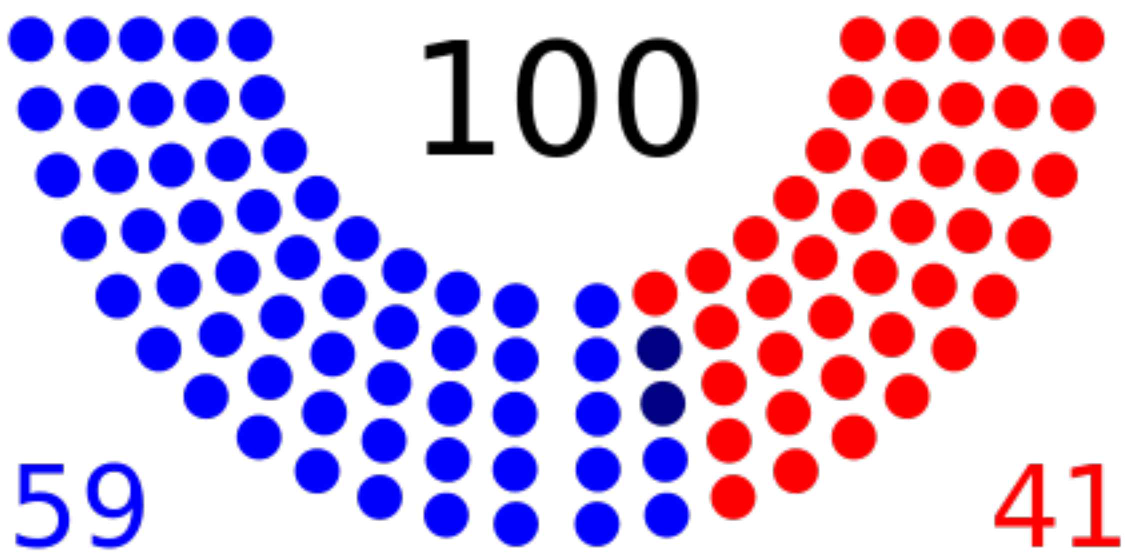


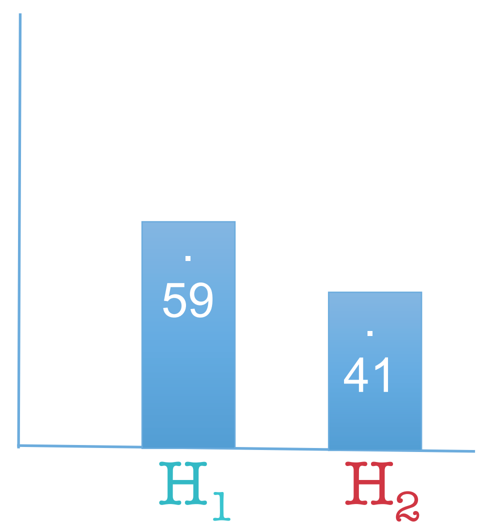
Prior: Initial belief

What's my best guess without any vote info?

I'd guess democrat since there are more of them.

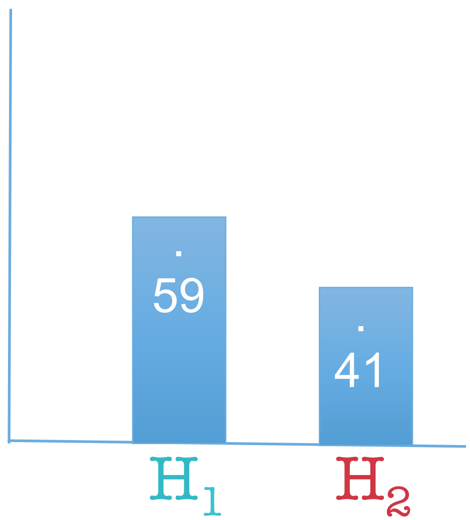
P(Democrat) = 0.59





Prior: Initial belief

P(Democrat) = 0.59

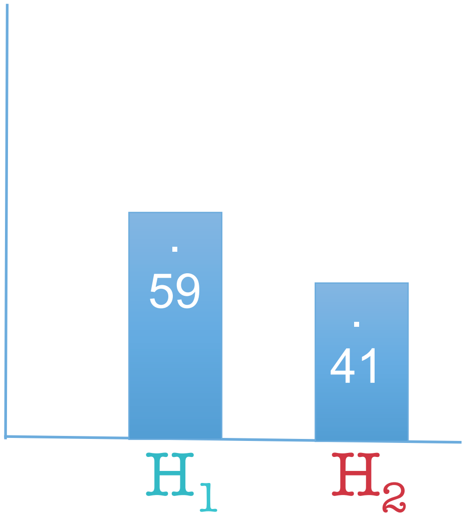


Prior: Initial belief

P(Democrat) = 0.59

New information (feature 1):

Voted YES on Net Neutrality



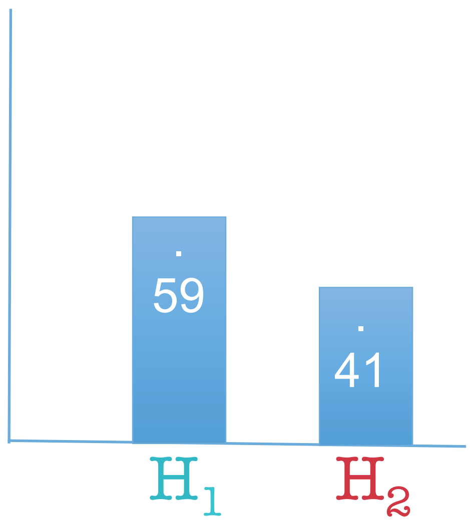
Prior: Initial belief

P(Democrat) = 0.59

New information (feature 1):

Voted YES on Net Neutrality

$$P(Dem|Y_{NN}) = \frac{P(Y_{NN}|Dem)\,P(Dem)}{P(Y_{NN})}$$



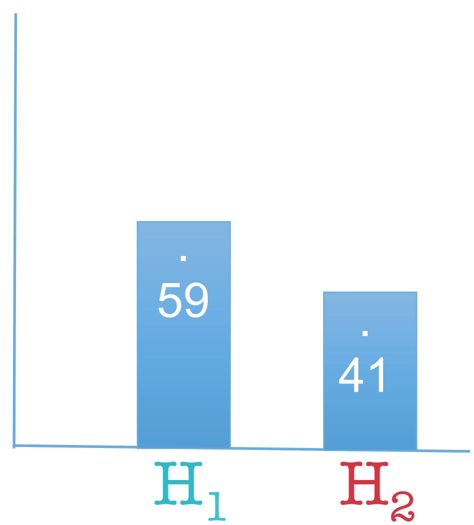
Prior: Initial belief

P(Democrat) = 0.59

New information (feature 1):

Voted YES on Net Neutrality

$$\underset{\text{posterior}}{P(Dem|Y_{NN})} = \frac{\overset{\text{likelihood}}{P(Y_{NN}|Dem)}\ \overset{\text{prior}}{P(Dem)}}{\underset{\text{evidence (normalization factor)}}{P(Y_{NN})}}$$



Prior: Initial belief

P(Democrat) = 0.59

$P(Y_{NN}|Dem)$

Prob. of voting yes on net neutrality if you're democrat

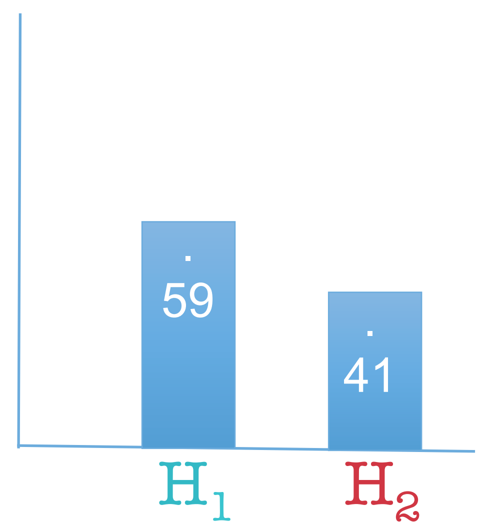
New information (feature 1):

Voted YES on Net Neutrality

posterior, likelihood, prior

$$P(Dem|Y_{NN}) = \frac{P(Y_{NN}|Dem)\,P(Dem)}{P(Y_{NN})}$$

evidence (normalization factor)



Prior: Initial belief

$P(\text{Democrat}) = 0.59$

New information (feature 1):

Voted YES on Net Neutrality

$P(Y_{NN}|\text{Dem})$

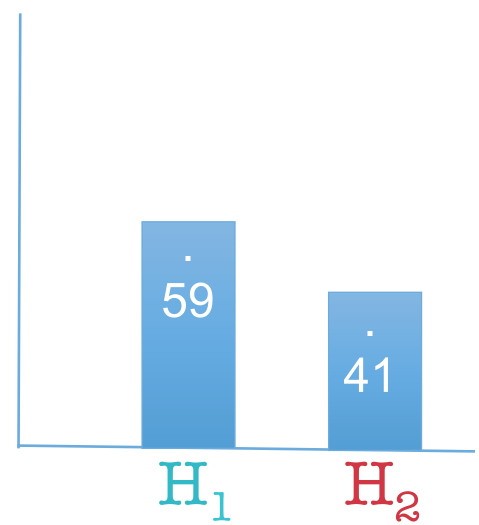
Prob. of voting yes on net neutrality if you're democrat

posterior, likelihood, prior

$$P(\text{Dem}|Y_{NN}) = \frac{P(Y_{NN}|\text{Dem})\,P(\text{Dem})}{P(Y_{NN})}$$

evidence (normalization factor)

$$P(\text{Rep}|Y_{NN}) = \frac{P(Y_{NN}|\text{Rep})\,P(\text{Rep})}{P(Y_{NN})}$$



Prior: Initial belief

P(Democrat) = 0.59

New information (feature 1):

Voted YES on Net Neutrality

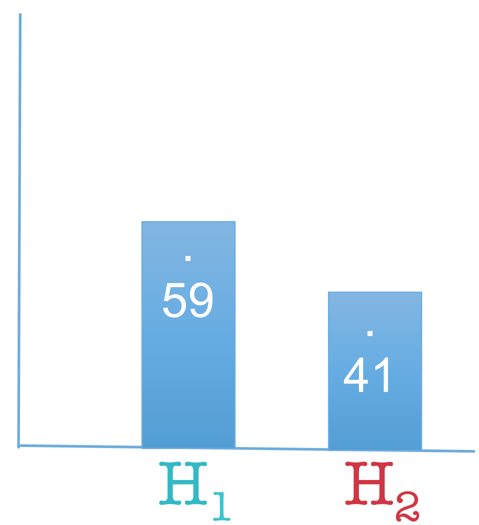
$P(Y_{NN}|Dem)$

Prob. of voting yes on net neutrality if you're democrat

$$\underset{\text{posterior}}{P(Dem|Y_{NN})} = \frac{\overset{\text{likelihood}}{P(Y_{NN}|Dem)}\ \overset{\text{prior}}{P(Dem)}}{\underset{\text{evidence (normalization factor)}}{P(Y_{NN})}}$$

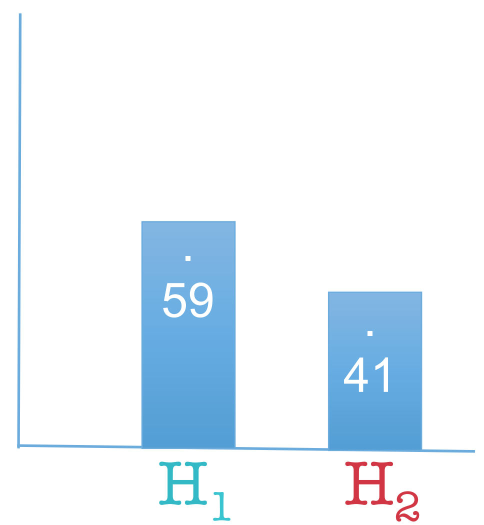
$P(Y_{NN}|Rep)$

$$P(Rep|Y_{NN}) = \frac{P(Y_{NN}|Rep)\ P(Rep)}{P(Y_{NN})}$$



$P(Y_{NN}|Dem)$
Prob. of voting yes
on net neutrality
if you're democrat

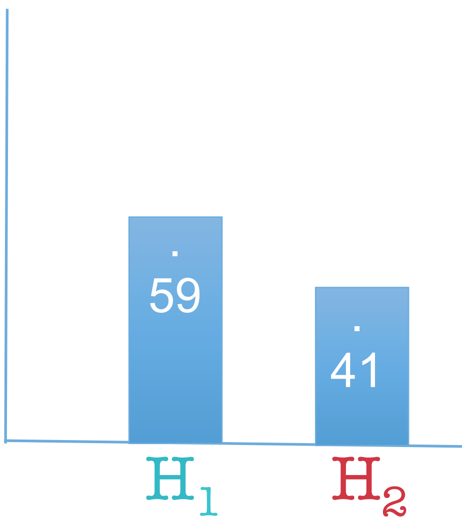
$P(Y_{NN}|Rep)$
Prob. of voting
yes
on net neutrality
if you're
republican



Training set has the answers!
We know Dem/Rep for each person, we know their votes!

$P(Y_{NN}|\text{Dem})$
Prob. of voting yes
on net neutrality
if you're democrat

$P(Y_{NN}|\text{Rep})$
Prob. of voting
yes
on net neutrality
if you're
republican



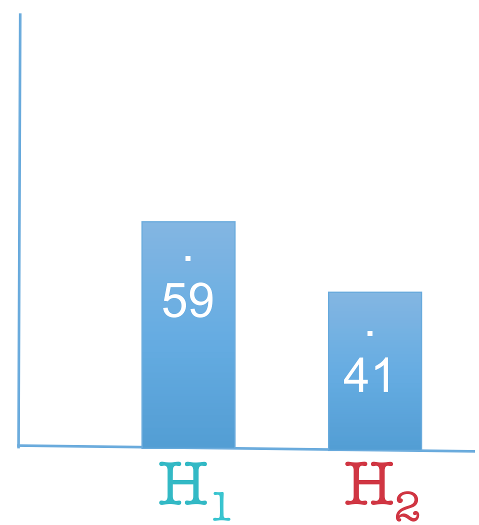
Training set has the answers!
We know Dem/Rep for each person, we know their votes!

$$P(Y_{NN}|Dem)$$

$$\approx \frac{\#\ \text{democrats that}\ Y_{NN}}{\#\ \text{all democrats}}$$

$$P(Y_{NN}|Rep)$$

$$\approx \frac{\#\ \text{republicans that}\ Y_{NN}}{\#\ \text{all republicans}}$$

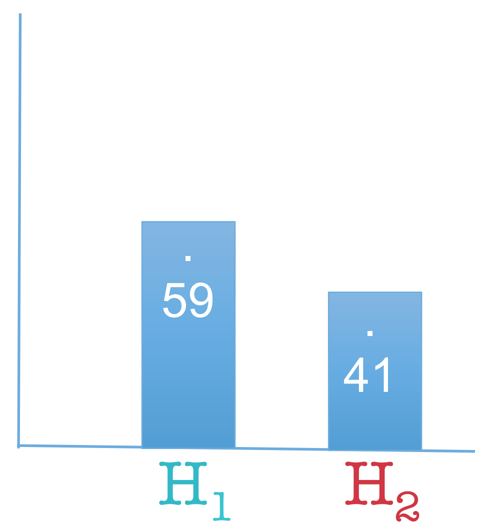


Training set has the answers!
We know Dem/Rep for each person, we know their votes!

$$P(Y_{NN}|Dem) \approx \frac{\text{\# democrats that } Y_{NN}}{\text{\# all democrats}}$$

$$P(Y_{NN}|Rep) \approx \frac{\text{\# republicans that } Y_{NN}}{\text{\# all republicans}}$$

For likelihoods of discrete data,
training/fitting means counting!
(and estimating likelihoods by dividing counts)

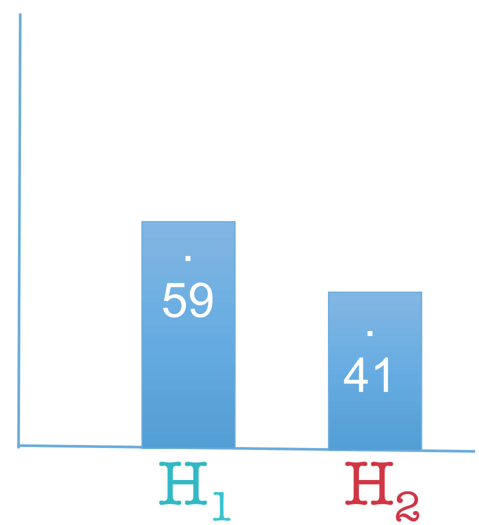


Training set has the answers!
We know Dem/Rep for each person, we know their votes!

$$P(Y_{NN}|\text{Dem}) \approx \frac{56}{59} = 0.949$$

$$P(Y_{NN}|\text{Rep}) \approx \frac{34}{41} = 0.829$$

For likelihoods of discrete data, training/fitting means counting!
(and estimating likelihoods by dividing counts)



Prior: Initial belief

P(Democrat) = 0.59

New information (feature 1):

Voted YES on Net Neutrality

likelihood prior

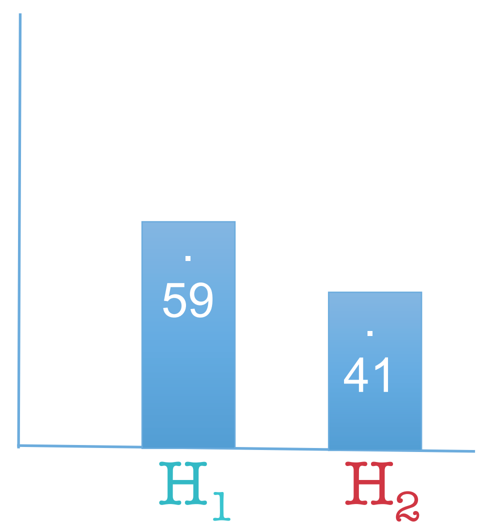
posterior

$$P(Dem|Y_{NN}) = \frac{P(Y_{NN}|Dem)\,P(Dem)}{P(Y_{NN})}$$

evidence

(normalization factor)

$$P(Rep|Y_{NN}) = \frac{P(Y_{NN}|Rep)\,P(Rep)}{P(Y_{NN})}$$



Prior: Initial belief

P(Democrat) = 0.59

New information (feature 1):

Voted YES on Net Neutrality

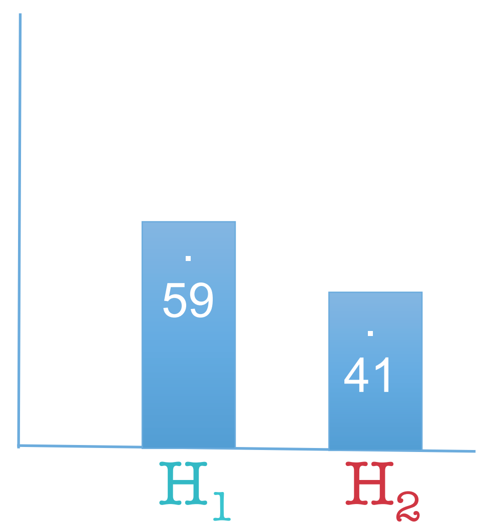
likelihood prior

posterior

$$P(Dem|Y_{NN}) = \frac{0.949 \quad * \quad 0.59}{P(Y_{NN})}$$

evidence
(normalization factor)

$$P(Rep|Y_{NN}) = \frac{0.829 \quad * \quad 0.41}{P(Y_{NN})}$$



Current belief

$P(\text{Democrat} \mid Y_{NN}) = 0.62$

New information (feature 1):

Voted YES on Net Neutrality

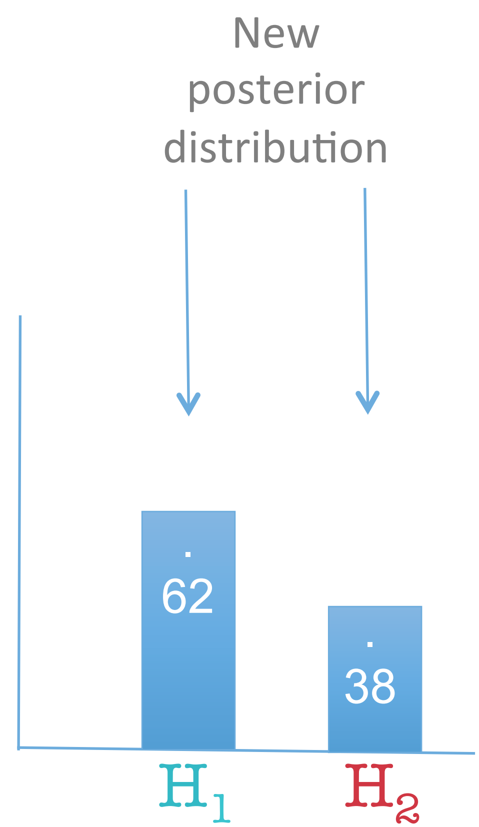
likelihood prior

posterior

$$P(\text{Dem} \mid Y_{NN}) = \frac{0.949 \;\; * \;\; 0.59}{P(Y_{NN})}$$

evidence
(normalization factor)

$$P(\text{Rep} \mid Y_{NN}) = \frac{0.829 \;\; * \;\; 0.41}{P(Y_{NN})}$$

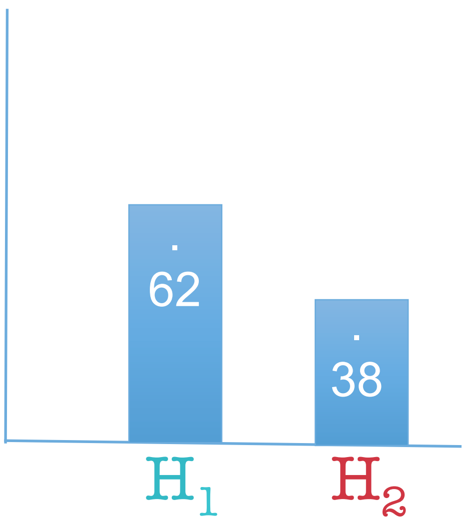


Current belief

$P(\text{Democrat} \mid Y_{NN}) = 0.62$

New information (feature 2):

Voted YES on Tax Cuts



Current belief

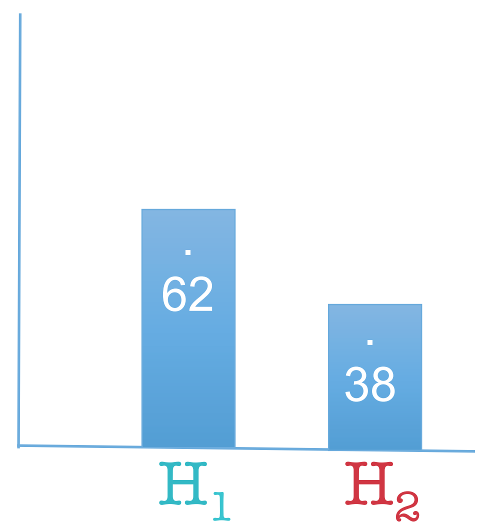
$P(\text{Democrat} \mid Y_{NN}) = 0.62$

New information (feature 2):

Voted YES on Tax Cuts

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}) = \frac{P(Y_{TC} \mid \text{Dem})\, P(\text{Dem} \mid Y_{NN})}{P(Y_{TC})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}) = \frac{P(Y_{TC} \mid \text{Rep})\, P(\text{Rep} \mid Y_{NN})}{P(Y_{TC})}$$



Current belief

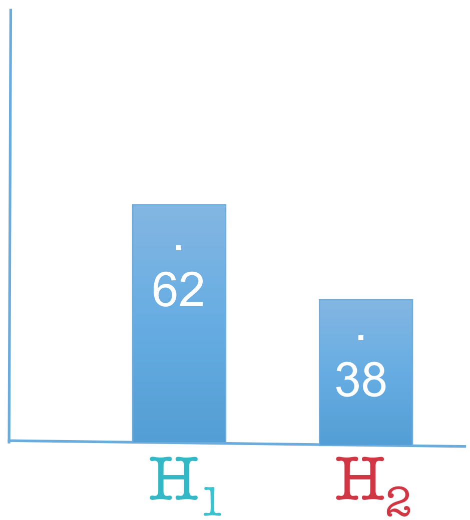
$P(\text{Democrat} \mid Y_{NN}) = 0.62$

$P(Y_{TC} \mid \text{Dem})$

$$\approx \frac{10}{59} = 0.169$$

$P(Y_{TC} \mid \text{Rep})$

$$\approx \frac{35}{41} = 0.854$$



Current belief

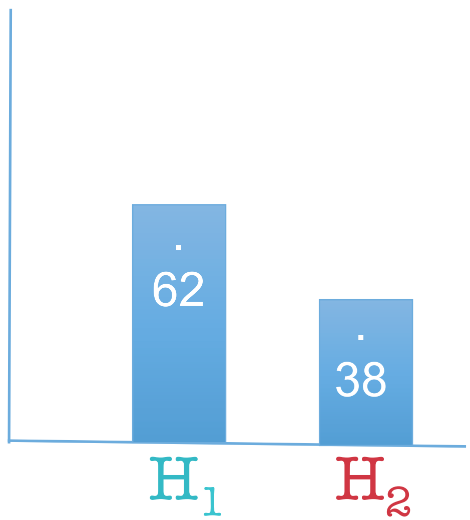
$P(\text{Democrat} \mid Y_{NN}) = 0.62$

New information (feature 2):

Voted YES on Tax Cuts

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}) = \frac{P(Y_{TC} \mid \text{Dem})\, P(\text{Dem} \mid Y_{NN})}{P(Y_{TC})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}) = \frac{P(Y_{TC} \mid \text{Rep})\, P(\text{Rep} \mid Y_{NN})}{P(Y_{TC})}$$



Current belief

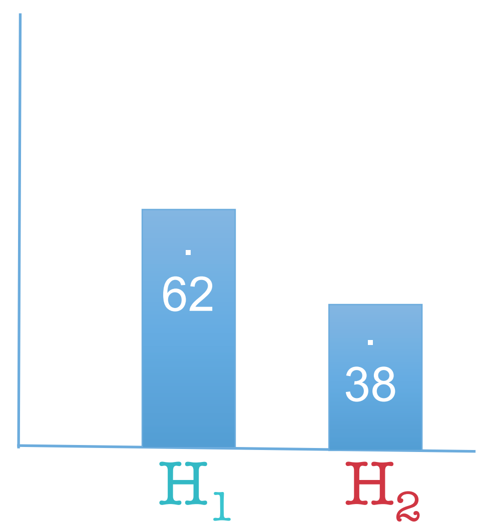
$P(\text{Democrat} \mid Y_{NN}) = 0.62$

New information (feature 2):

Voted YES on Tax Cuts

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}) = \frac{0.169 \quad * \quad 0.62}{P(Y_{TC})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}) = \frac{0.854 \quad * \quad 0.38}{P(Y_{TC})}$$



Current belief

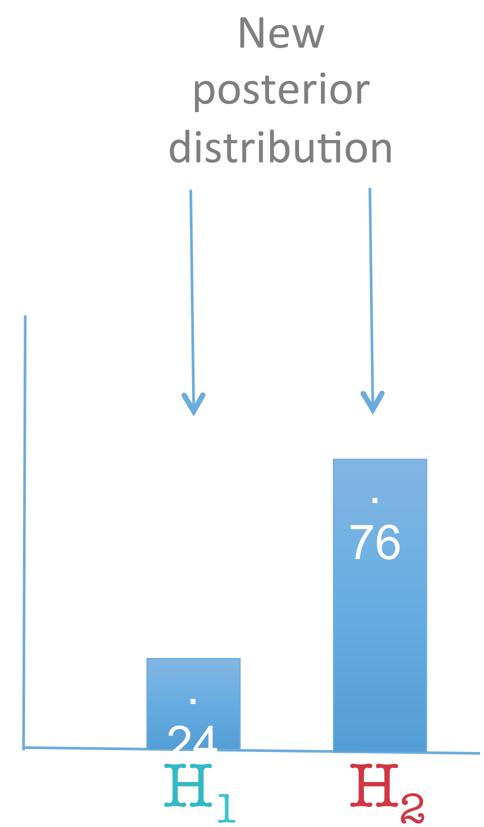
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}) = 0.24$

New information (feature 2):

Voted YES on Tax Cuts

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}) = \frac{0.169 \quad * \quad 0.62}{P(Y_{TC})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}) = \frac{0.854 \quad * \quad 0.38}{P(Y_{TC})}$$

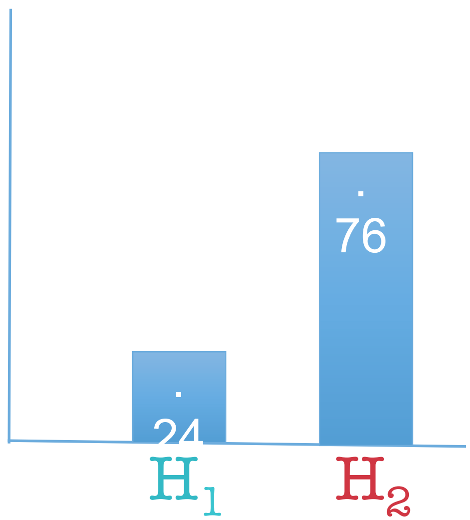


Current belief

$P(\text{Democrat} \mid Y_{NN}, Y_{TC}) = 0.24$

New information (feature 3):

Voted NO on License-free Guns



Current belief

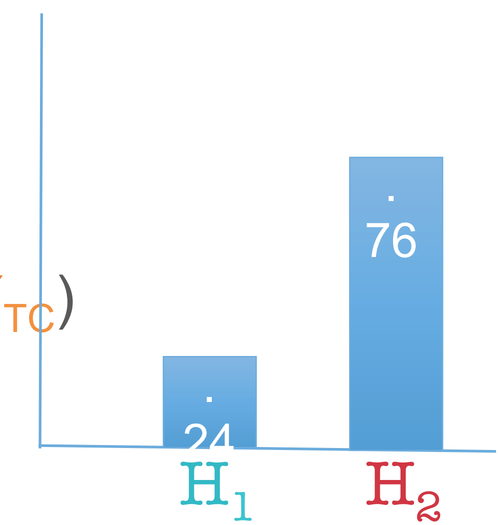
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}) = 0.24$

New information (feature 3):

Voted NO on License-free Guns

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{P(N_{LG} \mid \text{Dem})\, P(\text{Dem} \mid Y_{NN}, Y_{TC})}{P(N_{LG})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{P(N_{LG} \mid \text{Rep})\, P(\text{Rep} \mid Y_{NN}, Y_{TC})}{P(N_{LG})}$$



Current belief

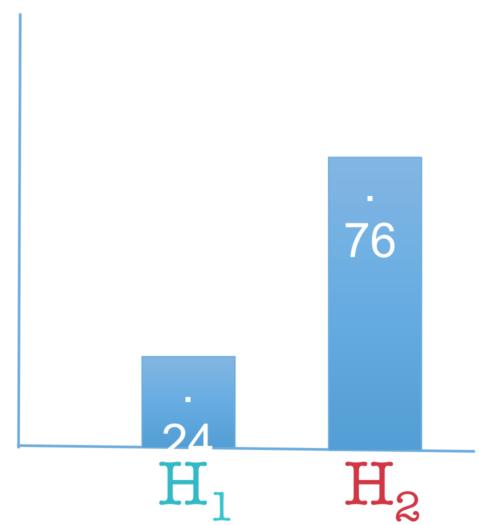
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}) = 0.24$

$P(N_{LG} \mid \text{Dem})$

$$\approx \frac{53}{59} = 0.898$$

$P(N_{LG} \mid \text{Rep})$

$$\approx \frac{23}{41} = 0.561$$



Current belief

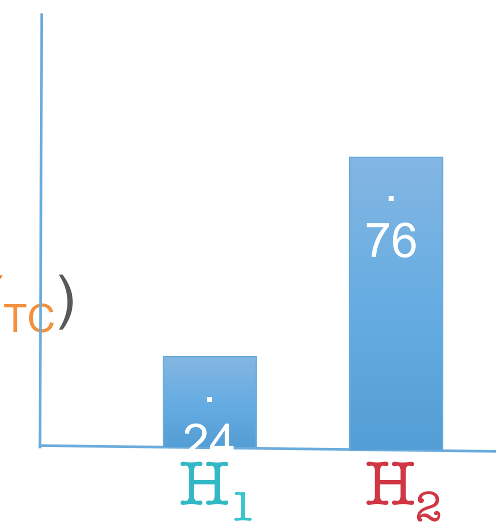
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}) = 0.24$

New information (feature 3):

Voted NO on License-free Guns

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{P(N_{LG} \mid \text{Dem})\, P(\text{Dem} \mid Y_{NN}, Y_{TC})}{P(N_{LG})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{P(N_{LG} \mid \text{Rep})\, P(\text{Rep} \mid Y_{NN}, Y_{TC})}{P(N_{LG})}$$



Current belief

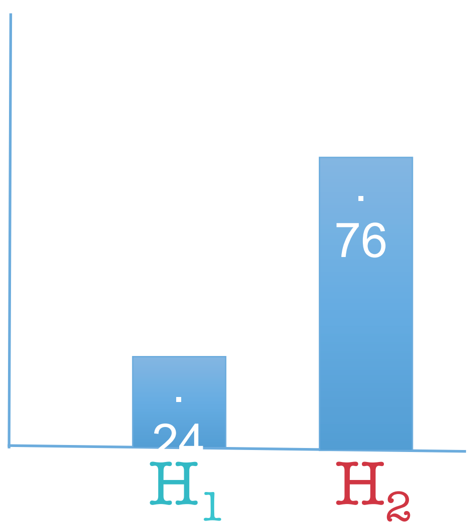
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}) = 0.24$

New information (feature 3):

Voted NO on License-free Guns

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{0.898 \quad * \quad 0.24}{P(N_{LG})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{0.561 \quad * \quad 0.76}{P(N_{LG})}$$



Current belief

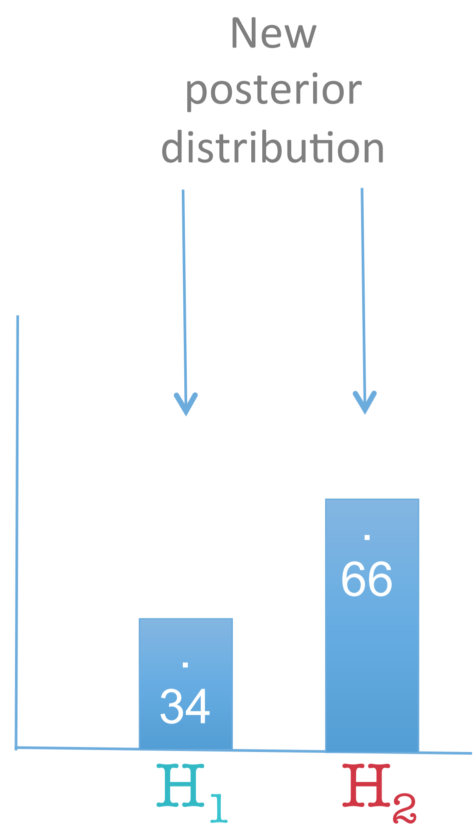
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}, N_{LG}) = 0.34$

New information (feature 3):

Voted NO on License-free Guns

$$P(\text{Dem} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{0.898 \quad * \quad 0.24}{P(N_{LG})}$$

$$P(\text{Rep} \mid Y_{NN}, Y_{TC}, N_{LG}) = \frac{0.561 \quad * \quad 0.76}{P(N_{LG})}$$



Current belief

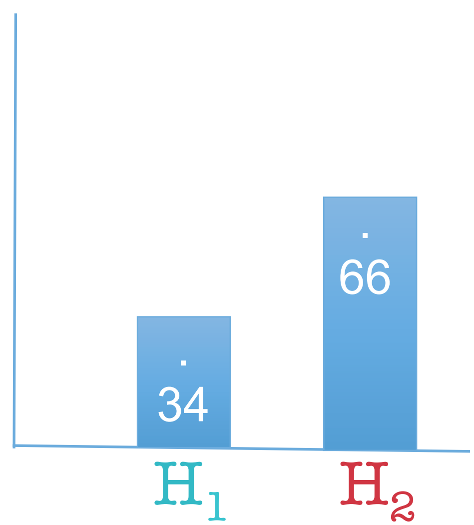
$P(\text{Democrat} \mid Y_{NN}, Y_{TC}, N_{LG}) = 0.34$

Classify this person that voted

Yes on Net Neutrality (Y_{NN}),

Yes on Tax Cuts (Y_{TC}),

No on License-free Guns (N_{LG})



Current belief

$P(\text{Democrat} \mid Y_{NN}, Y_{TC}, N_{LG}) = 0.34$

Classify this person that voted

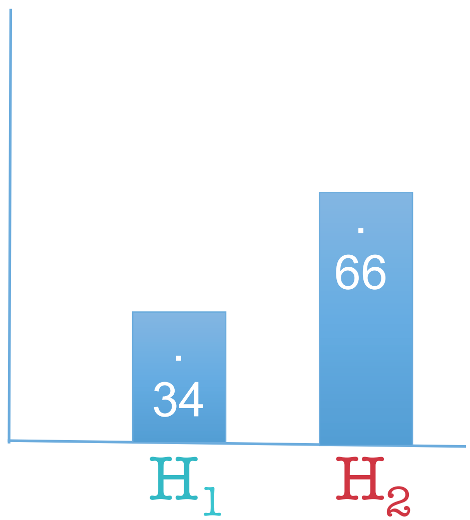
Yes on Net Neutrality (Y_{NN}),

Yes on Tax Cuts (Y_{TC}),

No on License-free Guns (N_{LG})

My strongest belief is in H_2,

I classify this person with the label Republican.



Naïve Bayes

Training:
Count and calculate the likelihood of each feature value for each class:

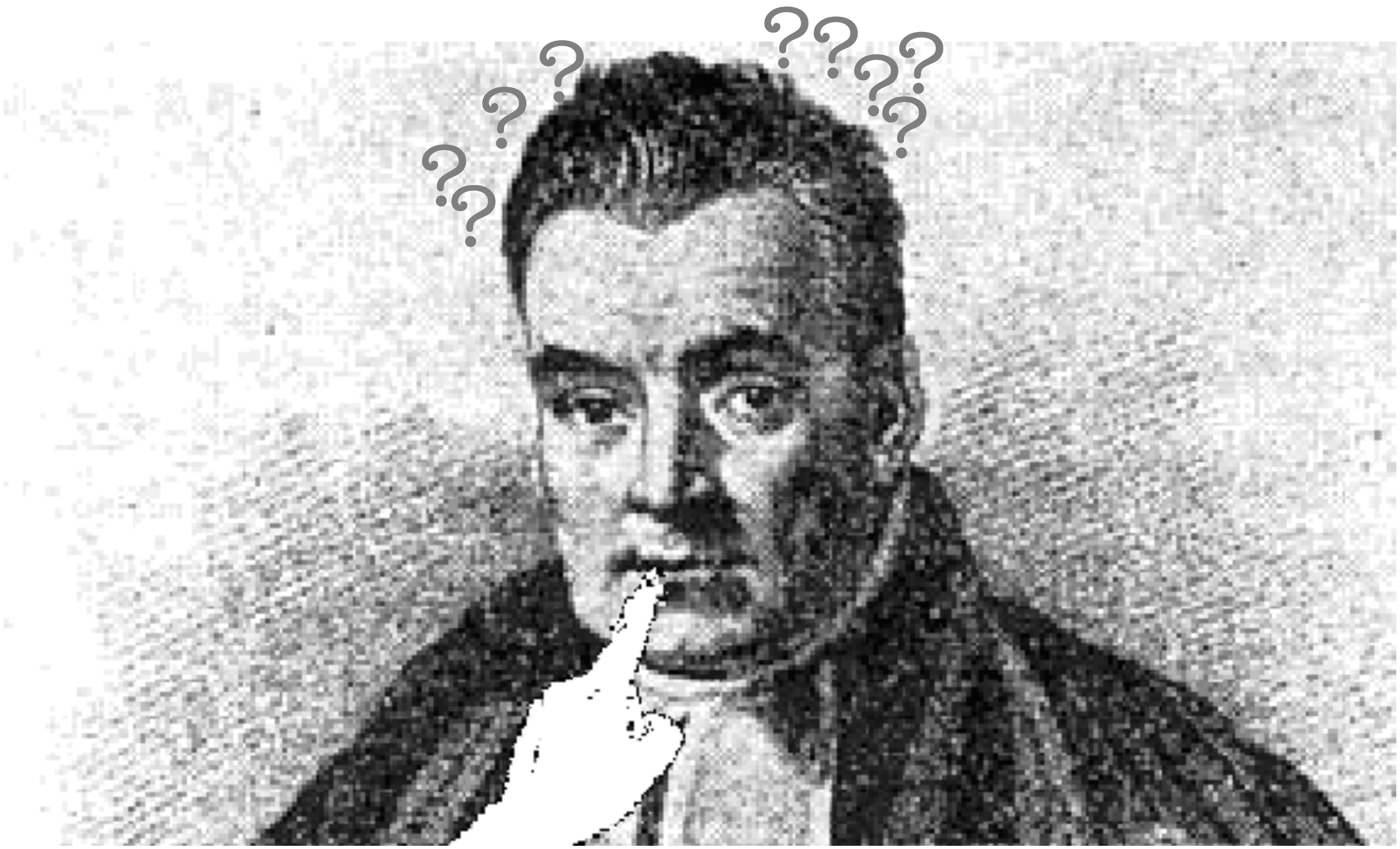
$P(Y_{NN}|Dem) \quad = 1 - P(N_{NN}|Dem)$
$P(Y_{NN}|Rep) \quad = 1 - P(N_{NN}|Rep)$

$P(Y_{TC}|Dem) \quad = 1 - P(N_{TC}|Dem)$
$P(Y_{TC}|Rep) \quad = 1 - P(N_{TC}|Rep)$

$P(Y_{LG}|Dem) \quad = 1 - P(N_{LG}|Dem)$
$P(Y_{LG}|Rep) \quad = 1 - P(N_{LG}|Rep)$

Prediction:
Use Bayes to update priors with the likelihoods,
Pick label with the highest posterior probability.

What was the naïve part?



We assumed each feature is independent

We assumed each feature is independent

Easier to see in a single update rather than sequential

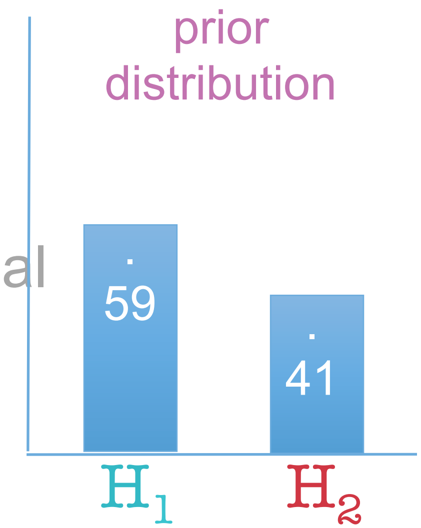
We assumed each feature is independent

Easier to see in a single update rather than sequential

Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

$$P(Dem|Y_{NN},Y_{TC},N_{LG}) = \frac{P(Y_{NN},Y_{TC},N_{LG}|Dem) \quad P(Dem)}{P(Y_{NN},Y_{TC},N_{LG})}$$

We assumed each feature is independent

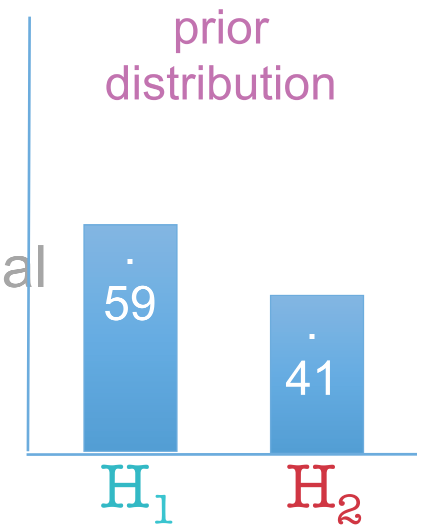


Easier to see in a single update rather than sequential

Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

$$\overset{\text{posterior}}{P(Dem|Y_{NN},Y_{TC},N_{LG})} = \frac{\overset{\text{likelihood}}{P(Y_{NN},Y_{TC},N_{LG}|Dem)} \; \overset{\text{prior}}{P(Dem)}}{P(Y_{NN},Y_{TC},N_{LG})}$$

We assumed each feature is independent



Easier to see in a single update rather than sequential

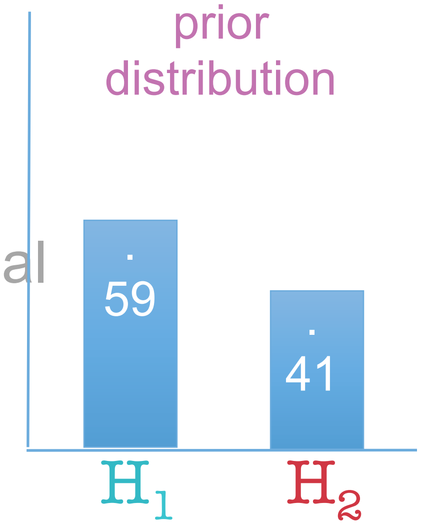
Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

$$\underbrace{P(Dem|Y_{NN},Y_{TC},N_{LG})}_{\text{posterior}} = \frac{\overbrace{P(Y_{NN},Y_{TC},N_{LG}|Dem)}^{\text{likelihood}} \; \overbrace{P(Dem)}^{\text{prior}}}{P(Y_{NN},Y_{TC},N_{LG})}$$

Independence Assumption:

$$P(Y_{NN},Y_{TC},N_{LG}|Dem) = P(Y_{NN}|Dem)\, P(Y_{TC}|Dem)\, P(N_{LG}|Dem)$$

We assumed each feature is independent



Easier to see in a single update rather than sequential

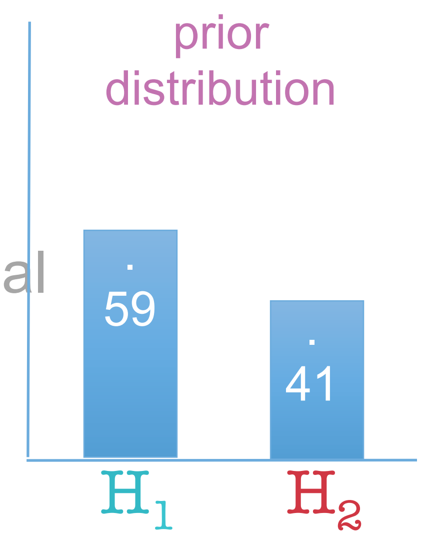
Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

$$\underbrace{P(Dem|Y_{NN},Y_{TC},N_{LG})}_{\text{posterior}} = \frac{\overbrace{P(Y_{NN},Y_{TC},N_{LG}|Dem)}^{\text{likelihood}}\ \overbrace{P(Dem)}^{\text{prior}}}{P(Y_{NN},Y_{TC},N_{LG})}$$

Independence Assumption:
$P(Y_{NN},Y_{TC},N_{LG}|Dem) = P(Y_{NN}|Dem)\ P(Y_{TC}|Dem)\ P(N_{LG}|Dem)$

Not even close in most cases!
Naïve Bayes still works well.

We assumed each feature is independent



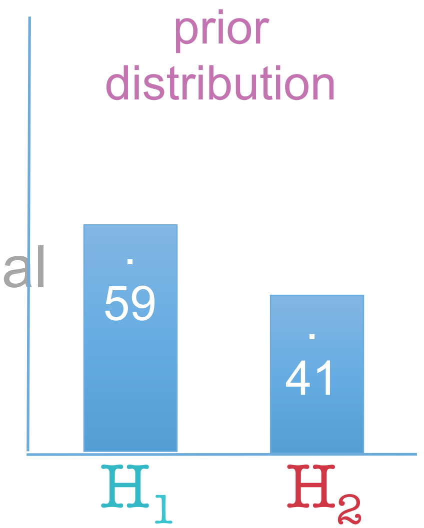
Easier to see in a single update rather than sequential

Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

likelihood — prior

posterior

$$P(Dem|Y_{NN},Y_{TC},N_{LG}) = \frac{P(Y_{NN}|Dem)\,P(Y_{TC}|Dem)\,P(N_{LG}|Dem)\,P(Dem)}{P(Y_{NN},Y_{TC},N_{LG})}$$

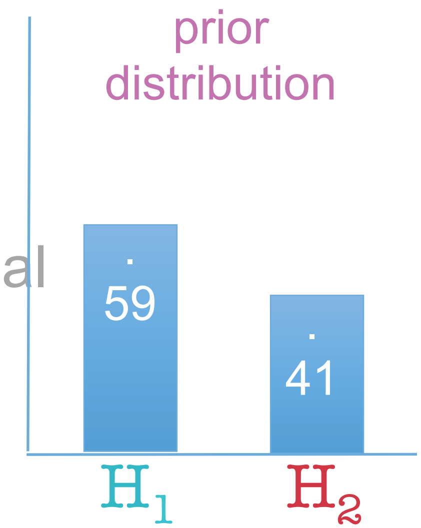


We assumed each feature is independent

Easier to see in a single update rather than sequential
Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

$$\underbrace{P(Dem \mid Y_{NN}, Y_{TC}, N_{LG})}_{\text{posterior}} = \frac{\overbrace{0.949 * 0.169 * 0.898}^{\text{likelihood}} \quad * \quad \overbrace{0.59}^{\text{prior}}}{P(Y_{NN}, Y_{TC}, N_{LG})}$$

We assumed each feature is independent



Easier to see in a single update rather than sequential

Prob. of this example having label Democrat,
given the values Yes, Yes and No
on the features NN, TC and LG

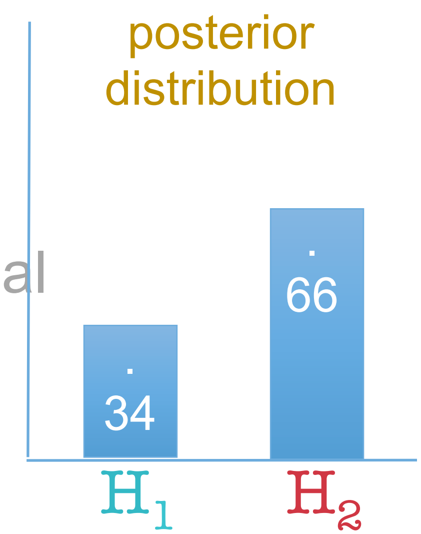
likelihood prior

posterior

$$P(Dem|Y_{NN},Y_{TC},N_{LG}) = \frac{0.949 * 0.169 * 0.898 \quad * \quad 0.59}{P(Y_{NN},Y_{TC},N_{LG})}$$

$$P(Rep|Y_{NN},Y_{TC},N_{LG}) = \frac{0.829 * 0.854 * 0.561 \quad * \quad 0.41}{P(Y_{NN},Y_{TC},N_{LG})}$$

We assumed each feature is independent



Easier to see in a single update rather than sequential

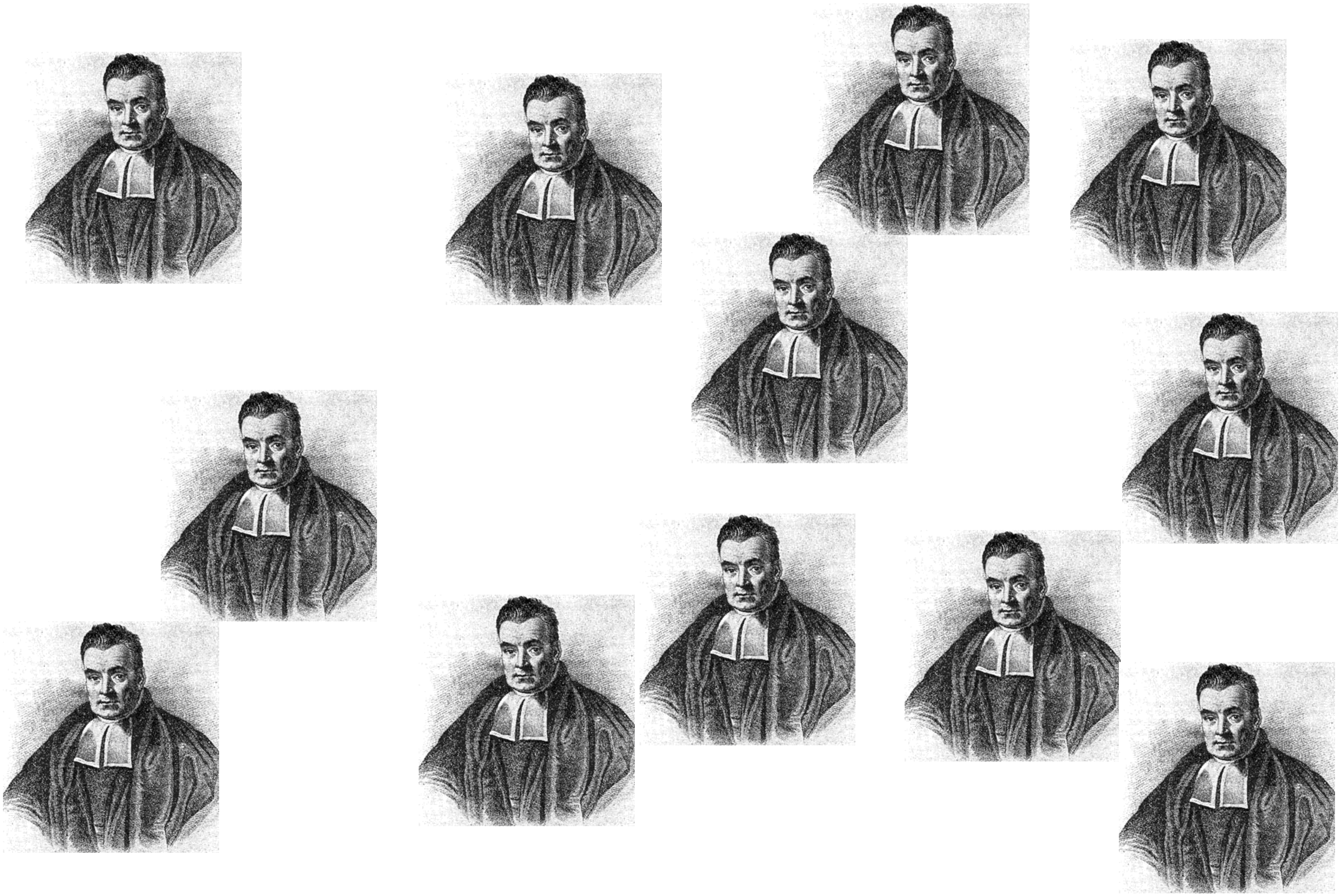
Prob. of this example having label Democrat, given the values Yes, Yes and No on the features NN, TC and LG

$$P(Dem|Y_{NN},Y_{TC},N_{LG}) = 0.34$$

$$P(Rep|Y_{NN},Y_{TC},N_{LG}) = 0.66$$

predict!

What about multiple classes?



Straightforward!

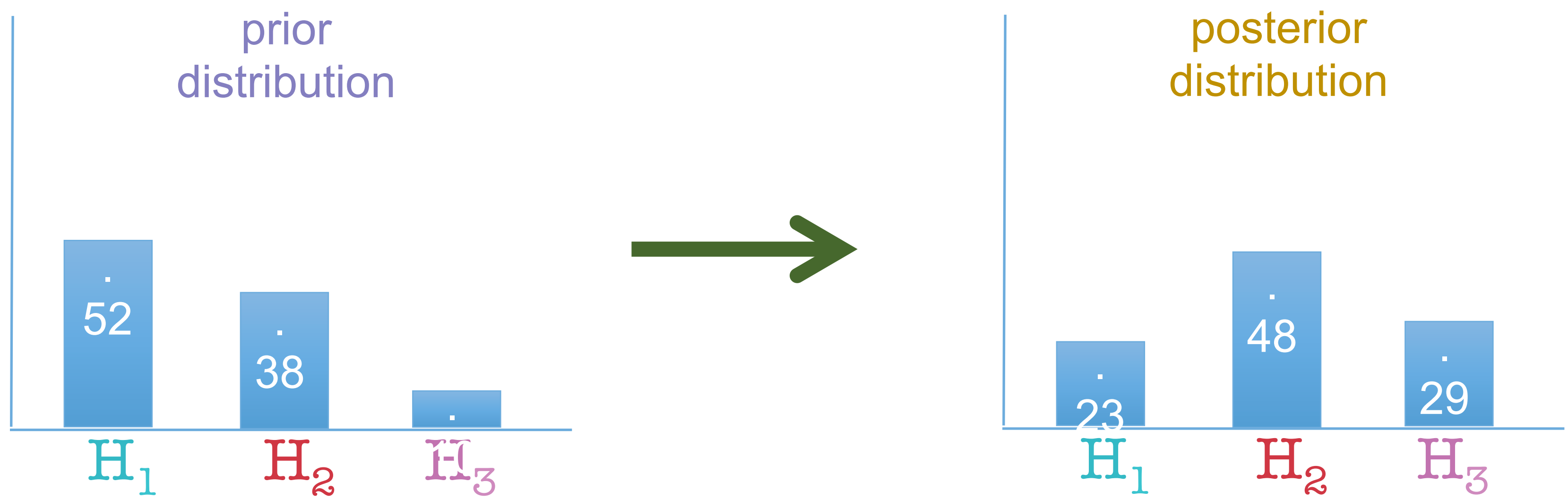
Update each hypothesis,
given the values Yes, Yes and No
on the features NN, TC and LG

$$P(Dem|Y_{NN},Y_{TC},N_{LG})$$
$$P(Rep|Y_{NN},Y_{TC},N_{LG})$$
$$P(Indep|Y_{NN},Y_{TC},N_{LG})$$

Straightforward!

Update each hypothesis,
given the values Yes, Yes and No
on the features NN, TC and LG

$$P(\text{Dem}|Y_{NN},Y_{TC},N_{LG})$$
$$P(\text{Rep}|Y_{NN},Y_{TC},N_{LG})$$
$$P(\text{Indep}|Y_{NN},Y_{TC},N_{LG})$$



Naïve Bayes: The Multinomial Approach (Text Classification example)

	Document	Class
	Hamburger NYC Hamburger	A
	Hamburger Hamburger LA	A
	Hamburger Cheeseburger	A
	Montreal Iceskate Hamburger	C
TEST	Hamburger Hamburger Hamburger Montreal Iceskate	??

1) Develop our likelihoods for ‘seeing’ each word in given the class:

$$P(\text{Hamburger} \mid A) = (5 + 1) / (8 + 6) = 3/7$$

smoothing

\# times ‘hamburger’ appeared in ‘A’ Docs

\# of unique words in all docs

\# of words in ‘A’ docs

Naïve Bayes: The Multinomial Approach

	Document	Class
	Hamburger NYC Hamburger	**A**
	Hamburger Hamburger LA	**A**
	Hamburger Cheeseburger	**A**
	Montreal Iceskate Hamburger	**C**
TEST	**Hamburger Hamburger Hamburger Montreal Iceskate**	**??**

1) Develop our likelihoods for 'seeing' each word in given the class:

P(Hamburger | A) = (5 + 1) / (8 + 6) = 3/ 7
P (Montreal| A) = (0 +1) / (8+6) = 1/14
P(Iceskate | A) = (0 +1) (8 +6) = 1/14

P(Hamburger | C)) = (1 + 1) / (3+6) = 2/9
P (Montreal| C) = (1+ 1) / (3+6) = 2 /9
P(Iceskate | C) = (1 + 1) /(3+6) = 2/ 9

P (A| test) ~ (3/7)^3*(1/14)*(1/14) * (3/4) = 0.0003
P (C | test) ~ (2/9)^3 *(2/9)*(2/9)* (1/4) = 0.0001

Naïve Bayes: The Bernouli Approach

	Document	Class
	Hamburger NYC Hamburger	**A**
	Hamburger Hamburger LA	**A**
	Hamburger Cheeseburger	**A**
	Montreal Iceskate Hamburger	**C**
TEST	**Hamburger Hamburger Hamburger Montreal Iceskate**	**??**

1) Develop our likelihoods for 'seeing' each word in given the class:

P(Hamburger | A) = (3+1) / (3+ 2) = 4/5

Constant = 2

of 'A' docs

smoothing

of 'A' docs
where 'American'
appears

Naïve Bayes: The Bernouli Approach

	Document	Class
	Hamburger NYC Hamburger	**A**
	Hamburger Hamburger LA	**A**
	Hamburger Cheeseburger	**A**
	Montreal Iceskate Hamburger	**C**
TEST	**Hamburger Hamburger Hamburger Montreal Iceskate**	**??**

1) Develop our likelihoods for 'seeing' each word in given the class:

P(Hamburger | A) = (3+1)/ (3+ 2) = 4/5
P (Montreal | A) = P(Iceskate | C) = (0+ 1) / (3+2) = 1/5
P(Cheeseburger| A) = P(NYC|A) =P(LA|A) (1+1) / (3+2) = 2/5

P(Hamburger|C) = P(Montreal)=P(Iceskate)=(1 + 1) / (1+2) = 2/3
P(NYC| C) = P(LA| C) =P(Cheeseburger| C) = (1+0)/(1+2) = 1/3

*P(A|test)=P(A)*P(Hamburger | A) *P (Montreal | A) *P(Iceskate | C) * (1-P(Cheeseburger| A))*(1-(P(NYC|A))(1-P(LA|A))*

P (A | test) ~(3/4)* (4/5)(1/5)(1/5)*(1-(2/5))*(1-(2/5))*(1-(2/5)) = .005
P (C |test) ~(1/4)* (2/3)(2/3)(2/3)*(1-(1/3))*(1-(1/3))*(1-(1/3)) = 0. 022

Naïve Bayes

Bernoulli : models the fraction of documents of class C that contain the word ‘w’ (ignores number of occurrences)

Vs.

Multinomial: models the fraction of *positions* in documents of class C that contain the word ‘w’ (keeps track of number of occurrences)

But why does Naïve Bayes work so well-
(Considering that it is Naïve) ?

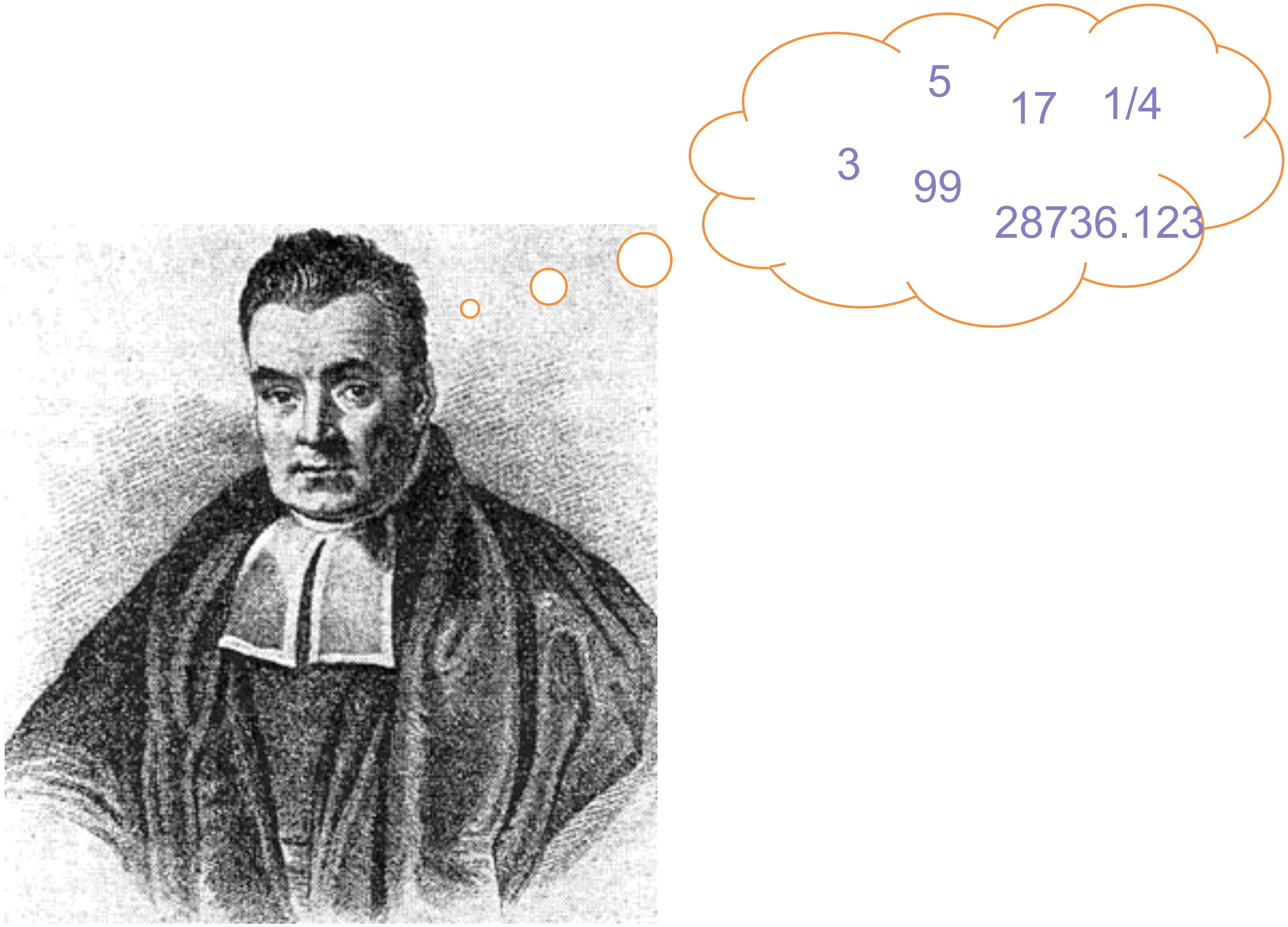
NB chooses among possible classes to find the class with the highest associated probability.

Naiveté doesn’t hurt, because correctness is based on classification, not exact predictions

Advantages of Naïve Bayes:

- Simple & Fast. Just doing a bunch of counts!
- Will converge quickly. Requires less training data
- Can handle sparse matrices
- Can handle multiple classes well

How about numeric features?



Naïve Bayes: The Gaussian Approach

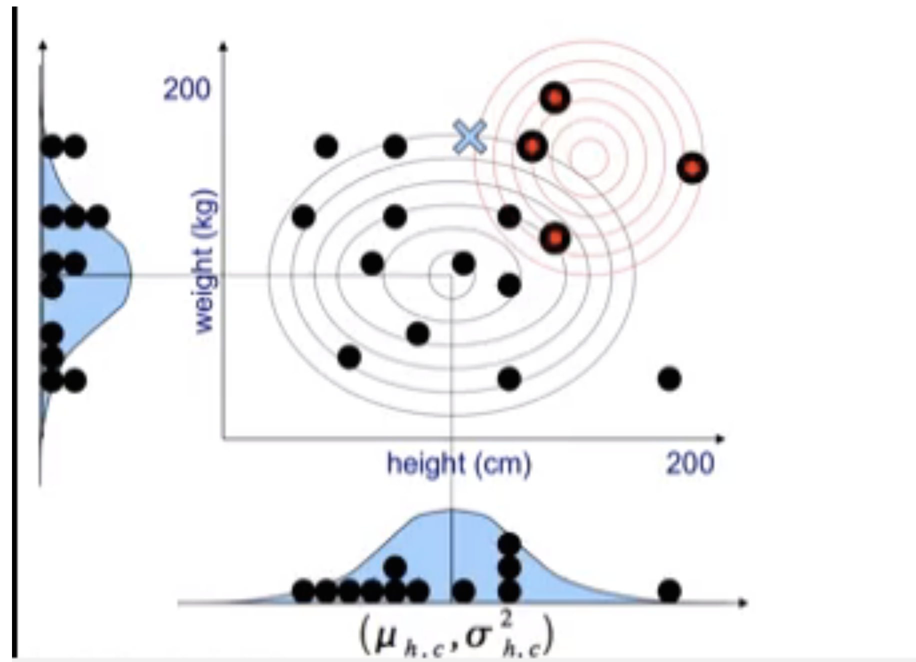
$$p(h_x|c)=\frac{1}{\sqrt{2\pi\sigma^2_{h,c}}}\exp-\frac{1}{2}\left(\frac{(h_x-\mu_{h,c})^2}{\sigma^2_{h,c}}\right)$$

$$p(w_x|c)=\frac{1}{\sqrt{2\pi\sigma^2_{w,c}}}\exp-\frac{1}{2}\left(\frac{(w_x-\mu_{w,c})^2}{\sigma^2_{w,c}}\right)$$

$$p(h_x|a)=\frac{1}{\sqrt{2\pi\sigma^2_{h,a}}}\exp-\frac{1}{2}\left(\frac{(h_x-\mu_{h,a})^2}{\sigma^2_{h,a}}\right)$$

$$p(w_x|a)=\frac{1}{\sqrt{2\pi\sigma^2_{w,a}}}\exp-\frac{1}{2}\left(\frac{(w_x-\mu_{w,a})^2}{\sigma^2_{w,a}}\right)$$

$$P(x|a)=p(h_x|a)p(w_x|a)$$



P(a|x) ~ P(x|a)*P(a)

Flavors of Bayes in `sklearn`:

Numeric Features:
Gaussian Naïve Bayes

Features that are 0 or 1 (and both matter):
Bernoulli Naïve Bayes

Features that are count-like (and only non-zero matters):
Multinomial Naïve Bayes

Which did we do?