Naive Bayes

Classification Problems with multiple classes

1. Given an article - predict which section of the newspaper (Current News, International, Arts, Sports, Fashion, etc)
2. Given a photo of a car number plate, identify which state it belongs to
3. Audio clip of a song, identify the genre

Classification Problems

1. All classification problems essentially equivalent to evaluating conditional probability
2. P(Yi |X) *i.e.* Given certain evidence X, what is the probability that this if from class Yi
3. Logistic Regression solves this by modelling the probabilistic relationship between X and Y
4. Such models are called Discriminative models

Naive Bayes Algorithm

1. Naive Bayes: Computes P(Yi | X) by using Bayes theorem and instead computes the inverse conditional probability P(X|Yi)
2. A simple classifier that performs surprisingly well on a large class of problems
3. This type of methods are called Generative Learning Models

US House of congress voting patterns

Class	V1	V2	V3	V4	V5	V6	V7
republican	n	y	n	y	y	y	n
republican	n	y	n	y	y	y	n
democrat	NA	y	y	NA	y	y	n

House Votes84 Dataset

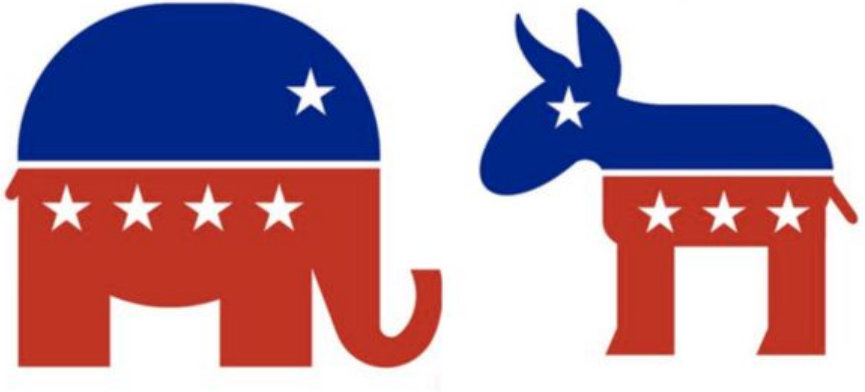
435 x 17

1. Class Name: 2 (democrat, republican)
2. handicapped-infants: 2 (y,n)
3. water-project-cost-sharing: 2 (y,n)
4. adoption-of-the-budget-resolution: 2 (y,n)
5. physician-fee-freeze: 2 (y,n)
6. el-salvador-aid: 2 (y,n)
7. religious-groups-in-schools: 2 (y,n)
8. anti-satellite-test-ban: 2 (y,n)
9. aid-to-nicaraguan-contras: 2 (y,n)
10. mx-missile: 2 (y,n)
11. immigration: 2 (y,n)
12. synfuels-corporation-cutback: 2 (y,n)
13. education-spending: 2 (y,n)
14. superfund-right-to-sue: 2 (y,n)
15. crime: 2 (y,n)
16. duty-free-exports: 2 (y,n)
17. export-administration-act-south-africa: 2 (y,n)

Republican or Democrat?

Given:
A **congressman's** voting pattern (v1=y, v2=n), what is the probability that this person is a **democrat**?

P(D|v1=y, v2=n)=?

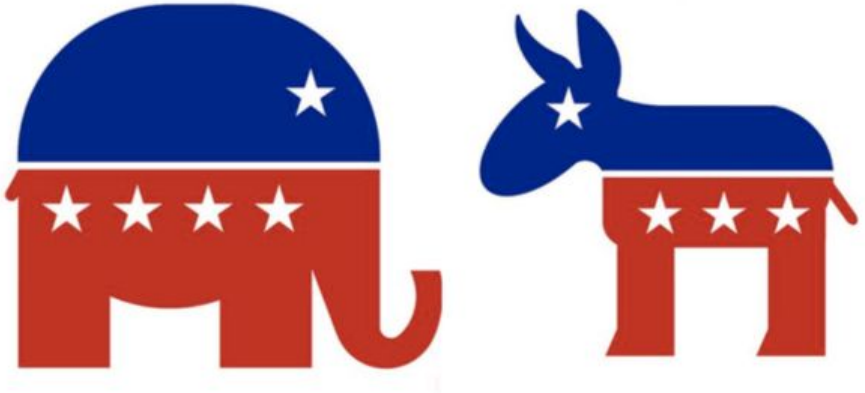


Republican or Democrat?

P(D | v1=y, v2=n)=?

Any guesses?

a) 0.2
b) 0.6
c) 0.9

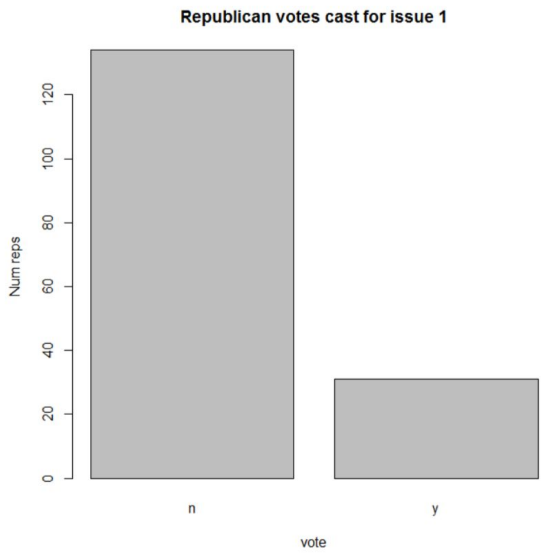


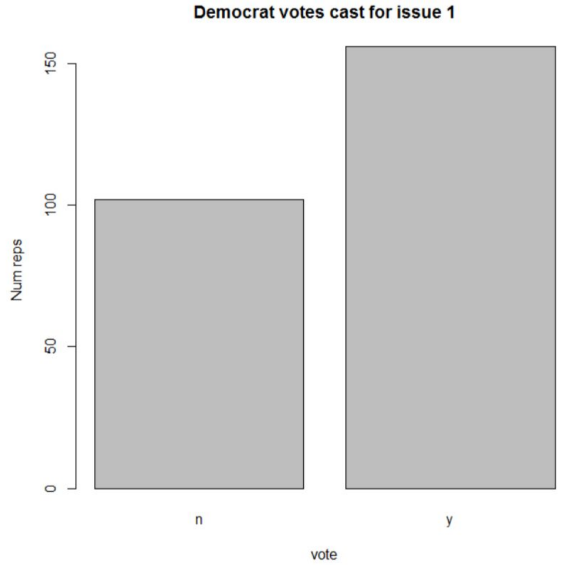
Prior Belief - Simplest Solution

1. The house has a majority of Democrats (168 Republicans, 267 Democrats)
2. Probability of a random person being Democrat is P(D) = 267/435 = 0.61
3. Can we do better?

Voting Patterns for V1 (Snapshot)

Handicapped infants. The vote failed to pass: 236 to 187



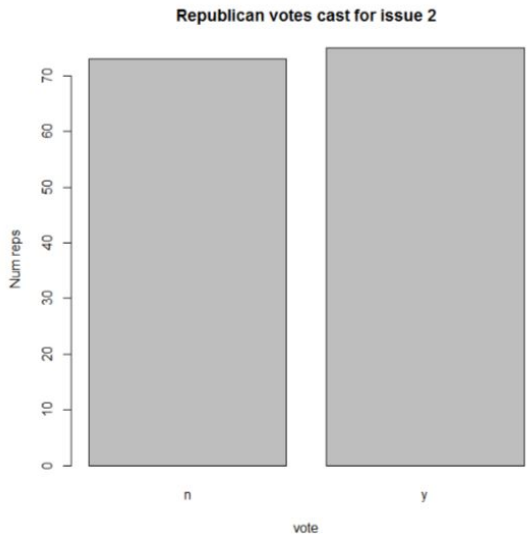


P (y | R) = 31/168

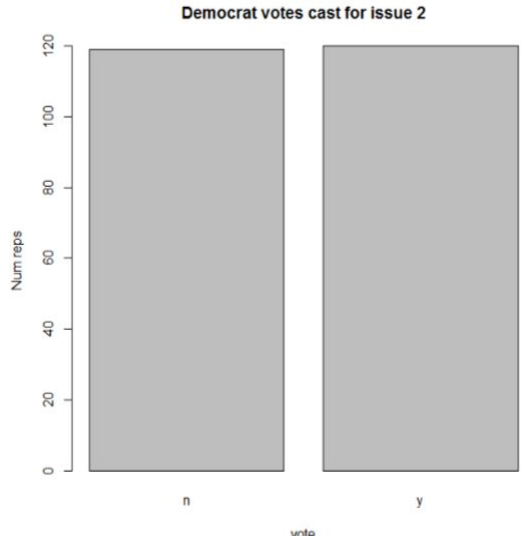
P (y | D) = 156/267

Voting Patterns for V2 (Snapshot)

Water-project-cost-sharing. The vote passed: 195 to 192



P (n | R) = 73/168



P (n | D) = 119/267

Bayes Theorem

$$P(A|B) = \frac{P(A) * P(B|A)}{P(B)}$$

$$P(D|\text{v1 = } y, \text{ v2=}n) = \frac{P(D) * P(\text{v1 = } y, \text{ v2=}n|D)}{P(\text{v1 = } y, \text{ v2=}n)}$$

Naive Bayes

Assumption: Conditional Probability of each feature given the class is independent of all other features

$$P(v1 = y, v2=n \mid D) = P(v1 = y \mid D) * P(v2 = n \mid D)$$

$$P(D \mid v1 = y, v2=n) = \frac{P(D) * P(v1 = y \mid D) * P(v2 = n \mid D)}{P(v1 = y, v2=n)}$$

Objective: What are we doing again?

Naive Bayes

We are trying to decide, **given** the voting pattern is that person a Democrat or Republican

Naive Bayes

We are trying to decide, **given** the voting pattern is that person a Democrat or Republican

$$P(D|v1 = y, v2=n) = \frac{P(D)*P(v1 = y|D)*P(v2 = n|D)}{P(v1 = y, v2=n)}$$

$$P(R|v1 = y, v2=n) = \frac{P(R) * P(v1 = y|R) * P(v2 = n|R)}{P(v1 = y, v2=n)}$$

Whichever probability is higher, we would classify the person into that party.
Note that the denominator is the same for both. So we need to focus only on numerator.

Naive Bayes - Key point in this slide?

We are trying to decide, **given** the voting pattern is that person a Democrat or Republican

$$P(D|v1 = y, v2=n) = \frac{P(D)*P(v1 = y|D)*P(v2 = n|D)}{P(v1 = y, v2=n)}$$

$$P(R|v1 = y, v2=n) = \frac{P(R) * P(v1 = y|R) * P(v2 = n|R)}{P(v1 = y, v2=n)}$$

Whichever probability is higher, we would classify the person into that party.
Note that the denominator is the same for both. So we need to focus only on numerator.

Naive Bayes

$$P(R|v1 = y, v2=n) = \frac{P(R) * P(v1 = y|R) * P(v2 = n|R)}{P(v1 = y, v2=n)}$$

Whichever probability is higher, we would classify the person into that party.
Note that the denominator is the same for both. **So we need to focus only on numerator.**

Naive Bayes - Recent Interview Question

Typically Naive Bayes is a fast algorithm

a) True

b) False

Naive Bayes - Recent Interview Question

Typically Naive Bayes is a fast algorithm

a) True
b) False

Ans) **a**

Naive Bayes (in class exercise - calculator)

$$P(D|v1 = y, v2=n) \propto P(D) * P(v1 = y|D) * P(v2 = n|D)$$

P(D) = 267/435 (267 Democrats among 435 Congressmen)

Naive Bayes

$$P(D|v1 = y, v2=n) \propto P(D) * P(v1 = y|D) * P(v2 = n|D)$$

P(D) = 267/435 (267 Democrats among 435 Congressmen)

$$P(D|v1 = y, v2=n) \propto \frac{267}{435} * \frac{156}{267} * \frac{119}{267} = 0.15$$

From voting pattern slide

$$P(R|v1 = y, v2=n) \propto \frac{168}{435} * \frac{31}{168} * \frac{73}{168} = 0.03$$

Naive Bayes

No parametric fit needed to compute the class!

Prior probability P(D) was computed from data

Individual conditional probabilities were evaluated, and using Bayes Relationship the final class probability was evaluated

```
from sklearn.naive_bayes import MultinomialNB
```

Naive Bayes - Assumption

The key assumption of independence of features, is almost never true

Still Naive Bayes does surprisingly well in a lot of situations

It works best when all the predictor variables are categorical variables

Very frequently used in text mining, character image analysis problems

SMS Classification

There is a 45% response rate for SMS messages (email has just 8%), but more importantly, the interaction with these texts is also significantly higher than other marketing tools. Classifying SMSes is extremely important. TrueCaller has hired you as a Machine Learning Engineer to make a SMS classification algorithm. How do we go about building this?

Dataset

30000 rows x 2 columns

3 Classes - ham, info, spam

	Label	Message
0	ham	oh how abt 2 days before Christmas
1	info	Welcome to OVATION HOLD R.No. 184, 114, 395, 3...
2	info	Thank you for using your ICICI bank CREDITcard...
3	ham	schedule a meeting with the entire team in the...
4	ham	Tommy is my brother
5	spam	OTP is 817453 for the txn of INR 8262.00 at SP...
6	ham	the meeting is scheduled by john
7	spam	Dear customer, We wish you a Merry Christmas. ...
8	spam	Delivered: Your package withPawzone Red 1.25 i...
9	info	The PNR for your Air India Flt 7I115 for PGH-B...
10	info	Bimal Auto Agency : Service of your car KA52C8...
11	info	Appointment with Dr Clayton in Pune on 2011-08...
12	info	Maha Veer Auto Agency : Service of your car KA...
13	spam	Dear AirAsia Customer, flight 5Q658 from RJA s...
14	info	Dear Guest, Thanks for choosing Forlini's Rest...
15	ham	I will indeed! What time?
16	info	Dear Guest, Thanks for choosing 2nd Avenue Del...

Steps

1. Preprocessing the text (Tokenization, lemmatization)
2. Split Train, Val, Test
3. Models built and evaluation

```
def message_to_words(raw_message):
#     letters_only = re.sub("[^a-zA-Z]", " ", raw_converse)
    # 1. Lower case & split
    words = raw_message.lower().split()
    # 2. Convert stop words to a set
    stops = set(stopwords.words("english"))
    # 3. Remove stop words
    meaningful_words = [w for w in words if not w in stops]
    # 4. Join the words back into one string
    return(b" ".join(meaningful_words))
```

Naive Bayes

Pretty High accuracy and recall!

```
# Naive Bayes Accuracy and recall

accuracy = metrics.accuracy_score(test.y,preds_NB)
recall = metrics.recall_score(test.y,preds_NB, average = 'macro')
print(accuracy, recall)
```

```
(0.9709999999999997, 0.9679800139378042)
```

Comparison with other models

Here is the link to the code and the dataset
https://drive.google.com/drive/u/2/folders/1aJDQqvcS3NtRKgvyTScTpjOG6PBGdVnq?ogsrc=32

Algorithms	Accuracy
Naive Bayes	0.97
MLP (Epoch = 15)	0.99
CNN 1D (Epoch = 10)	0.34
LSTM (Epoch = 1)	0.48

Thank you!