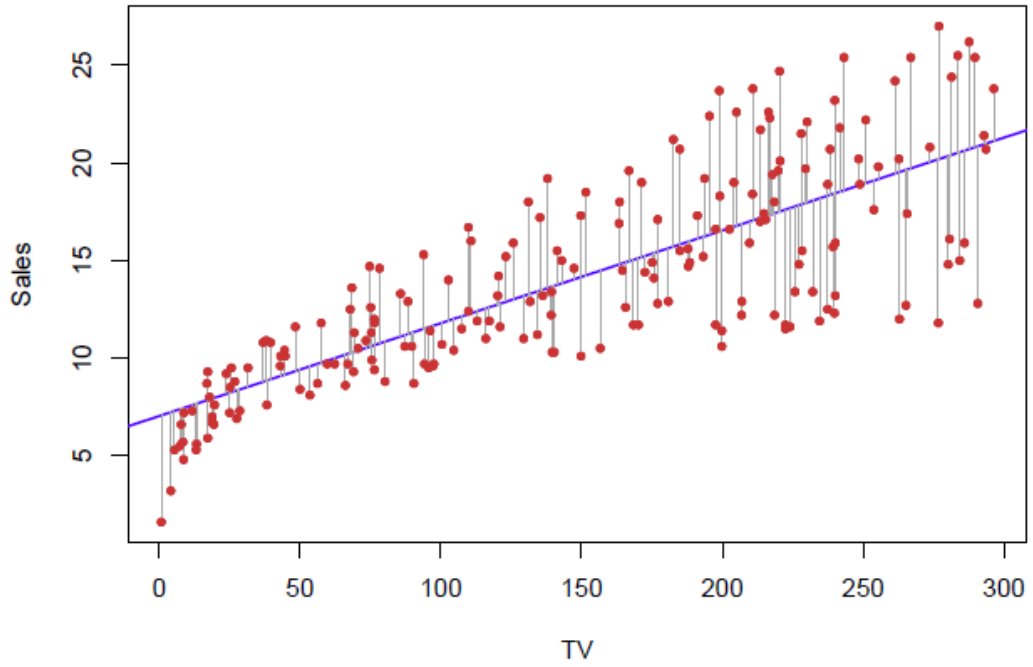
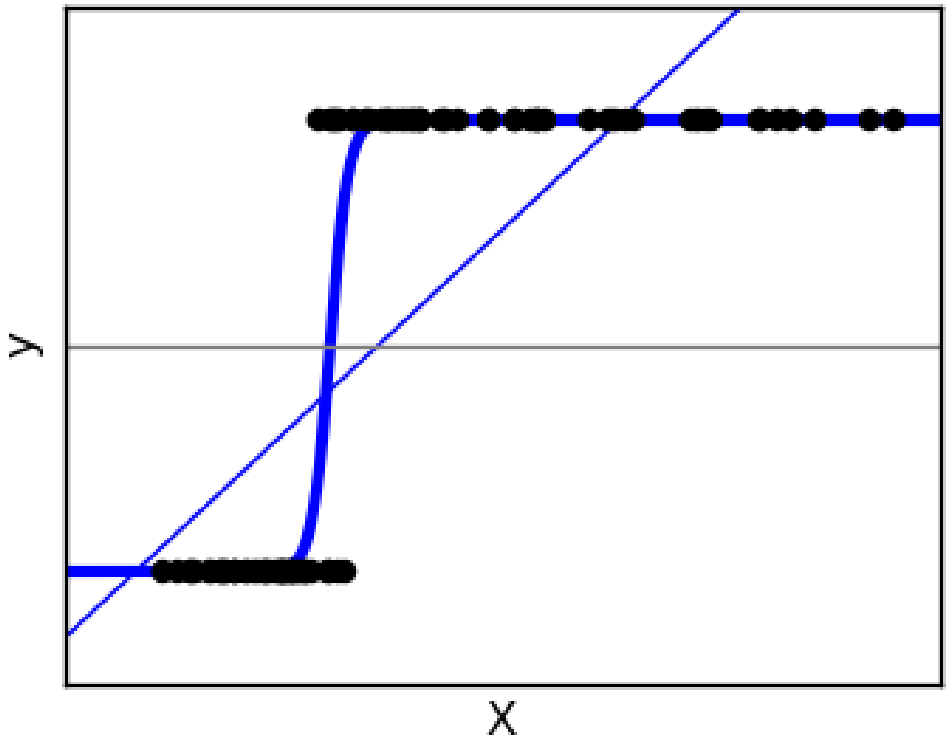
Naïve Bayes Algorithm

Praphul Chandra

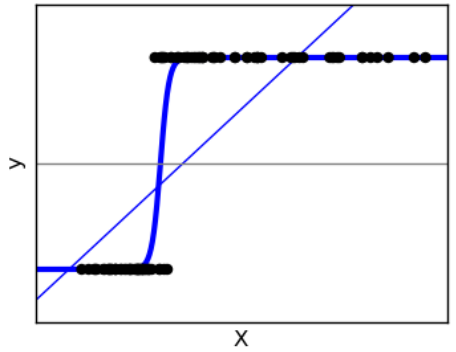
Supervised Learning as Conditional Probability





Classification as Conditional Probability

- All classification problems are essentially equivalent to evaluating conditional probability
  - $P(Y_i \mid X)$, *i.e.,* given certain evidence X, what is the probability that this is from class Y_i
  - Logistic Regression solves this problem by modelling the probabilistic relationship between X and Y (sigmoid function, linear in X, etc.) directly



- Naïve Bayes:
  - Computes $P(Y_i \mid X)$ using Bayes theorem
    - Leverage computation of the inverse conditional probability $P(X \mid Y_i)$ --- this might be easier!
  - A simple classifier that performs surprisingly well on a large class of problems

Classification Problems with multiple classes

- Given an article – predict which section of the news paper (Current News, International, Arts, Sports, Fashion, etc.) it is supposed to go to
- Given a photo of a car number plate, identify which state it belongs to
- Audio clip of a song, identify the genre

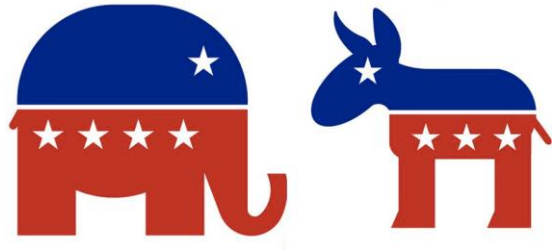
US House of Congress voting patterns

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)

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
democrat	n	y	y	n	NA	y	n
democrat	y	y	y	n	y	y	n
democrat	n	y	y	n	y	y	n
democrat	n	y	n	y	y	y	n
republican	n	y	n	y	y	y	n
republican	n	y	n	y	y	y	n
democrat	y	y	y	n	n	n	y
republican	n	y	n	y	y	n	n
republican	n	y	n	y	y	y	n
democrat	n	y	y	n	n	n	y
democrat	y	y	y	n	n	y	y
republican	n	y	n	y	y	y	n
republican	n	y	n	y	y	y	n
democrat	y	n	y	n	n	y	n
democrat	y	NA	y	n	n	n	y
republican	n	y	n	y	y	y	n
democrat	y	y	y	n	n	n	y

House Votes 1984 Dataset: Voting patterns of Members of Congress.
A data frame with 435 observations on 17 variables. 168 Republicans, 267 Democrats

Republican or Democrat?



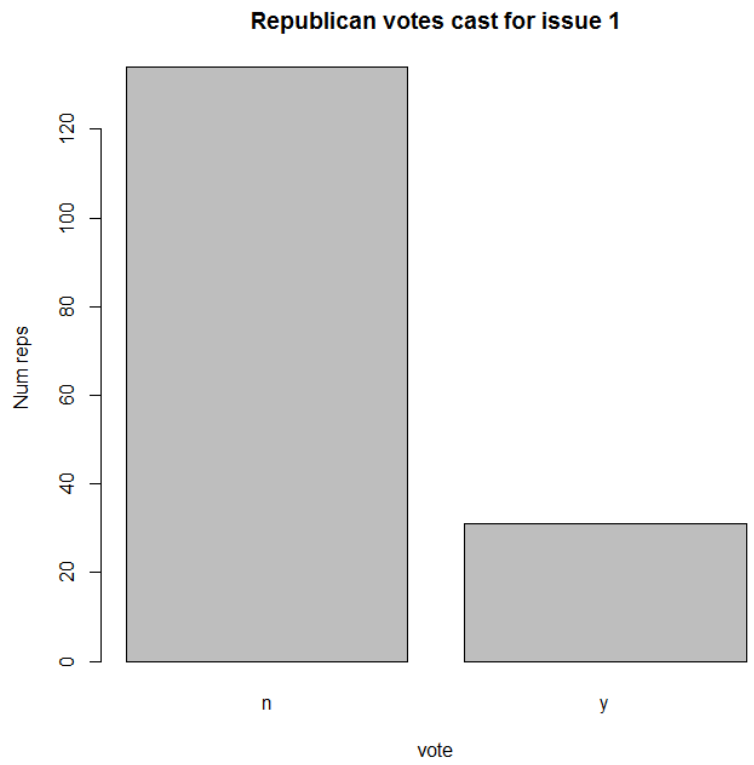
- Given:
  - A congressman's voting pattern (v1 = *y*, v2 = *n*), what is the probability that this person is a democrat?
- Predict
  - Whether a congressman is a republican or a democrat

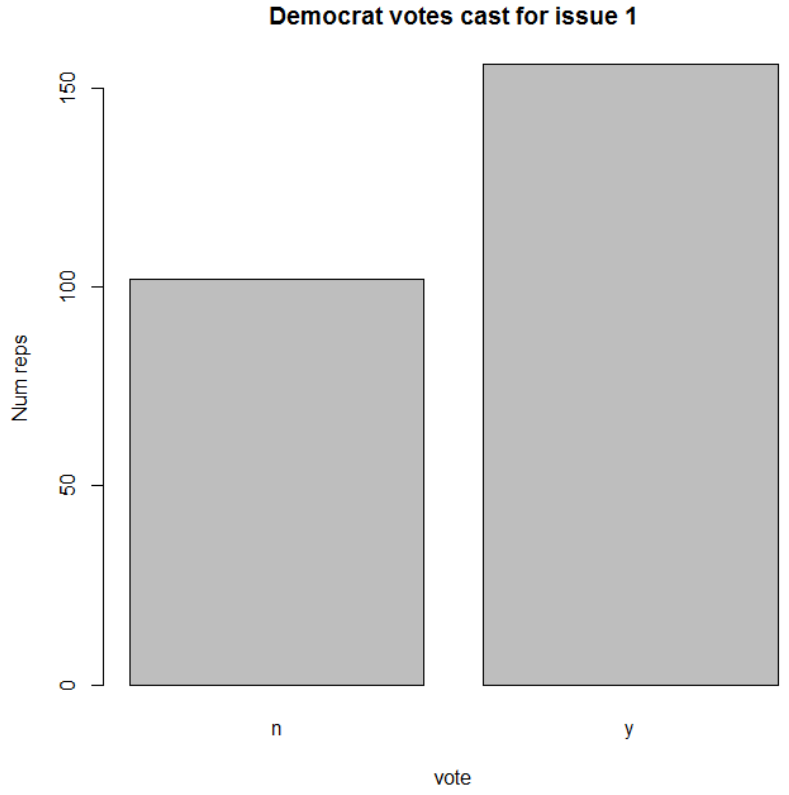
- In terms of probability
  - P(D | v1 = *y*, v2 = *n*) = ?

- Simplest (Naïve) Solution : Prior Belief
  - The house has a majority of Democrats (168 Republicans, 267 Democrats)
  - Probability of a random person being Democrat is P(D) = 267/435 = 0.61
  - Can we do better?

Voting Patterns for V1

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



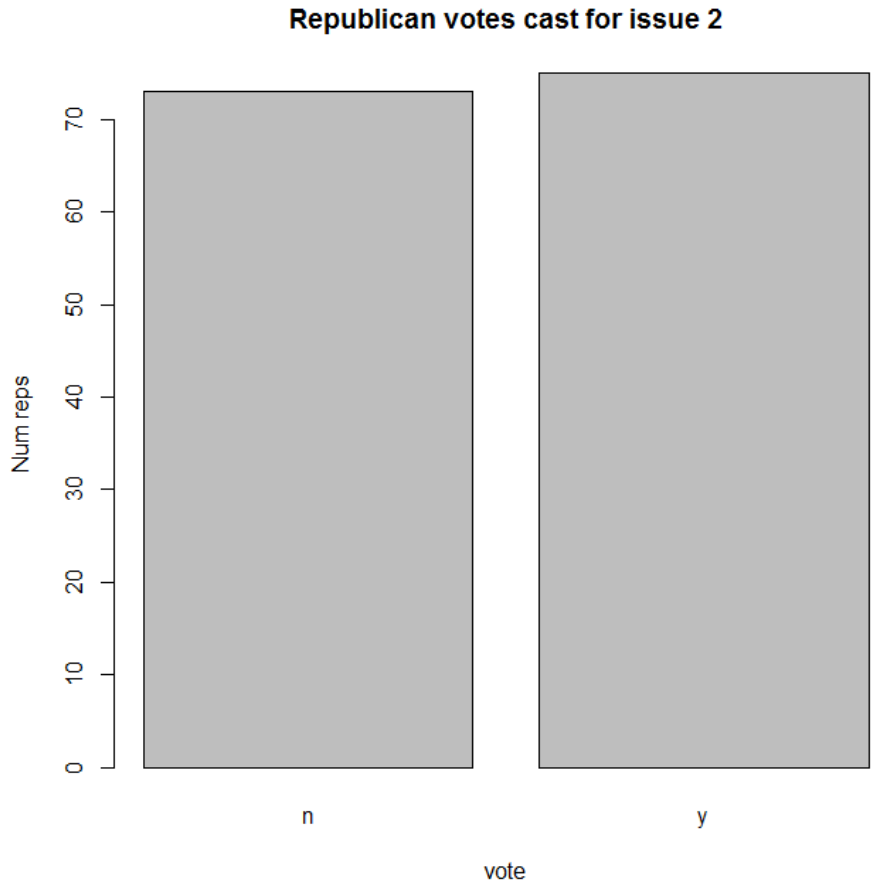


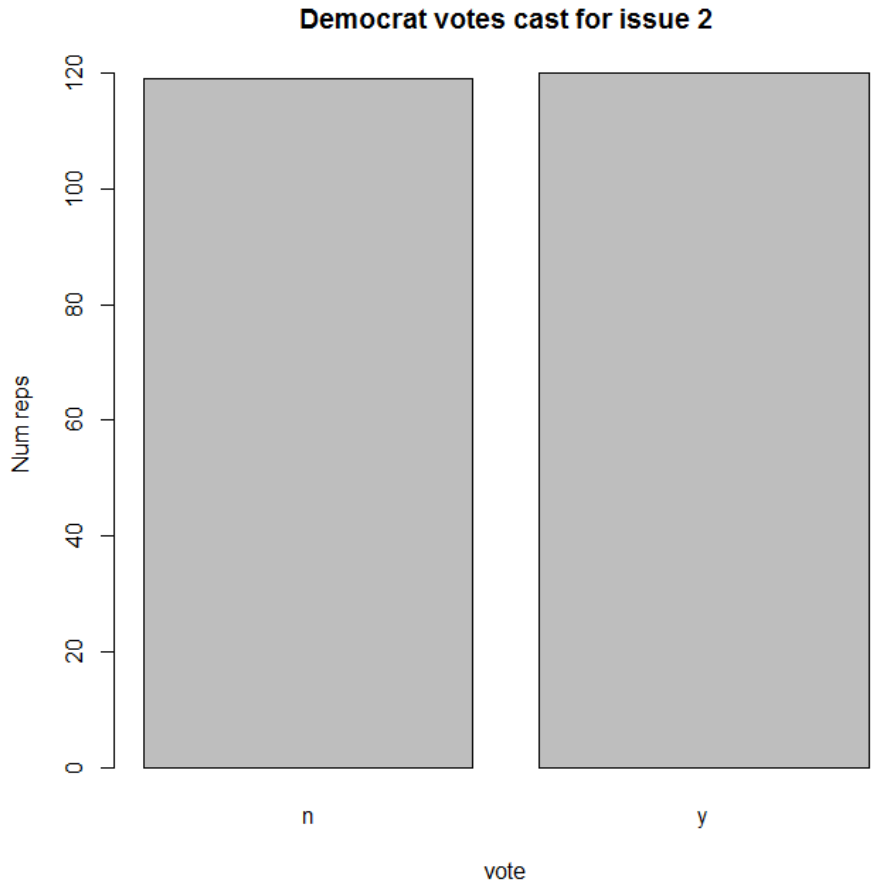
P(y | R) = 31/168

P(y | D) = 156/267

Voting Patterns for V2

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





P(n | R) = 73/168

P(n | D) = 119/267

Bayes Theorem

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

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

Naïve Bayes

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

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

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

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

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

- Whichever probability is higher, we would classify the person into that party.
  - We are trying to decide, given the voting pattern, is that person a Democrat or Republican.
  - Note that the denominator is the same for both. So we need to focus only on numerator.

Naïve Bayes Example

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

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

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

From voting pattern slide

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

Since the conditional probability for being Democrat is higher, he is likely to be Democrat.

Naïve Bayes Assumption

- The key assumption of independence of features, is almost never true
- Still Naïve 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

Q?

Praphul Chandra