# NAÏVE BAYES

Lara J. Martin (she/they)
TA: Aydin Ayanzadeh (he)

11/21/2023

CMSC 671

By the end of class today, you will be able to:

- Identify potential ethical issues from using probability models
- Recognize the components needed to setup a naïve bayes classifier

*Modified from slides by Dr. Chris Callison-Burch*

## Which equation is Bayes' Rule?

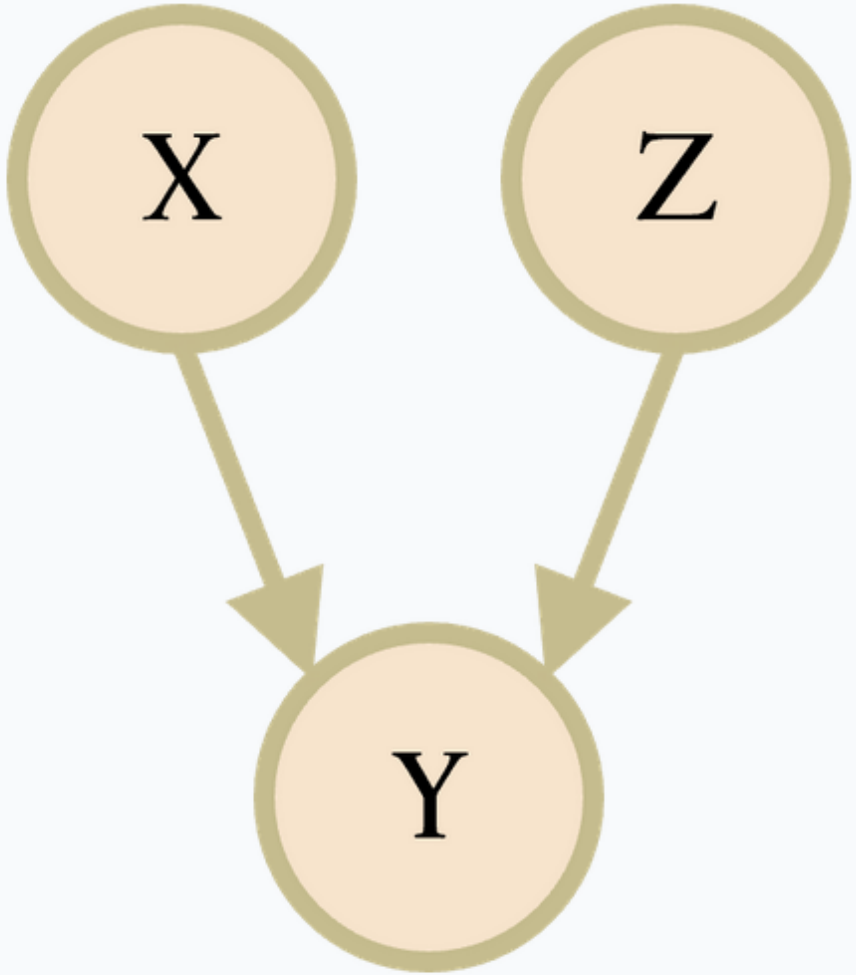
$P(H \mid E) = (P(E \mid H)P(H))/(P(E))$

$P(H,E) = P(E)P(H)$

$P(E)P(H|E) = P(E \mid H)P(H)$

$P(E) = (P(E \mid H)P(H))/(P(H|E))$

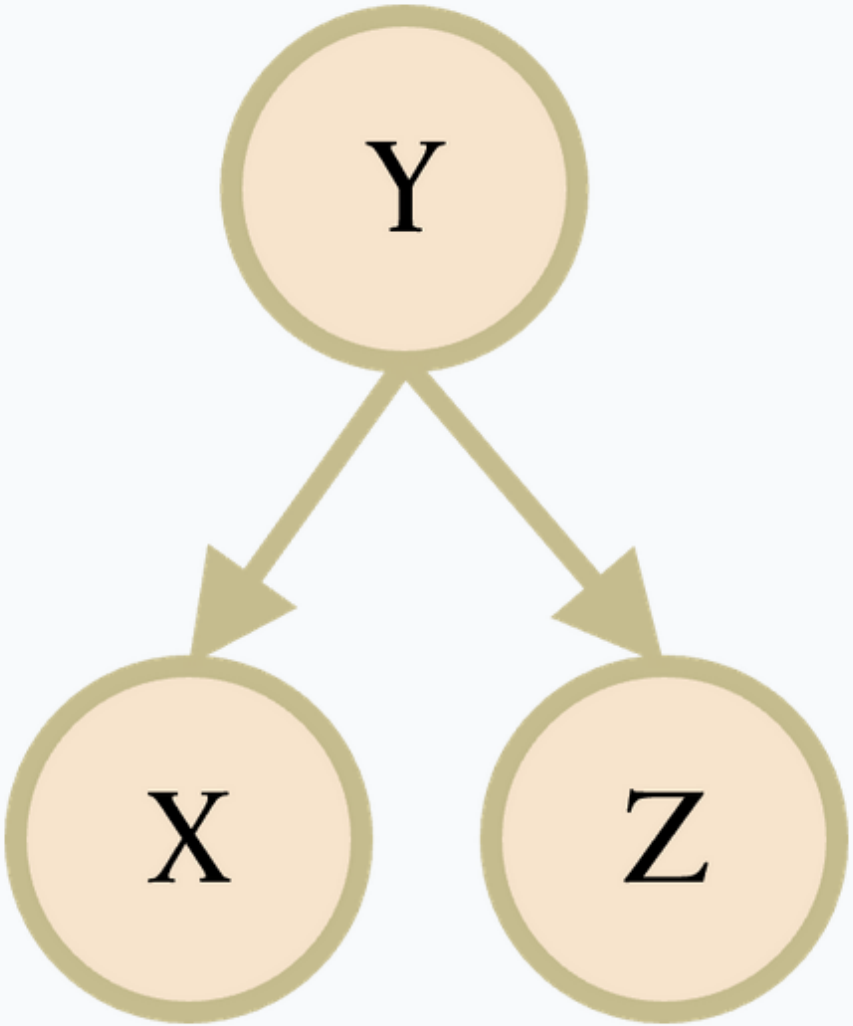
## In this Bayes' Net: If Y is given, are X & Z independent?




Yes 75%

No 25%

## In this Bayes' Net: If Y is given, are X & Z independent?




Yes

33%

No

67%

# FINAL PROJECT ADVISORS

Aydin

- Demographics & Obesity
- Blood Cell Image Classification
- Text Analysis & Summarization
- Fraud Detection

Lara

- Cybersecurity
- Smart Home Energy Optimization
- Flappy Bird RL
- Lunar Lander RL
- GPT Inconsistencies

# PROJECT MILESTONE 2

- Make progress on the suggested next steps that the class staff and your classmates left for you on your project proposal. – Will return by end of day tomorrow
- Read at least 3 papers related to your proposed project.
- Finish collecting any datasets that you need and make sure they are in a format that you can work with.
- Outline a more definite plan on the types of methods you want to use. What will your system look like?
  - If your project involves calling external libraries, do some quick experiments to see whether the methods you want to use are feasible.

# ETHICS OF PROBABILITY

# EXPLAINABILITY AND INTERPRETABILITY

- How clear is our agent's decision making? Is it transparent or is it a black box?
- Can we make changes to the algorithm to make its decisions more explainable?
- Can we develop tools that make the algorithm's decisions easier to interpret?

# INEQUALITY

- Who has access to this AI agent?
  - Could this create new inequality between groups that have access and do not have access?
- Is this system reinforcing existing structures that create inequality?
  - If yes, is there regulation for this technology that can prevent this?

# JOB DISPLACEMENT

- Will this algorithm displace human workers?
  - If yes, is there a plan in place to help those displaced workers?

- Will this algorithm/agent create new jobs? Who will benefit?

# NAÏVE BAYES

# BAYES' NETS
# INFERENCE BY ENUMERATION

- Requires lots of tedious calculations

$$P(d \mid e) = \alpha \Sigma_{ABC} P(a, b, c, d, e)$$

$$= \alpha \Sigma_{ABC} P(a)\, P(b \mid a)\, P(c \mid a)\, P(d \mid b, c)\, P(e \mid c)$$

- Remember: Exact inference in Bayes' Nets is NP-hard
- Gets unruly with too many variables

# NAÏVE BAYES ASSUMPTION

- Assume all features are independent effects of the label
- "**conditional independence assumption** that the probabilities $P(f_i|c)$ are independent given the class c and hence can be 'naively' multiplied as follows:

$$P(f_1, f_2, ...., f_n|c) = P(f_1|c) \cdot P(f_2|c) \cdot ... \cdot P(f_n|c)"$$

# CLASSIFICATION: SPAM OR NOT

Dear Sir.
First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. …


TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.
99 MILLION EMAIL ADDRESSES FOR ONLY $99


Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power

# SETTING UP A NAÏVE BAYES SPAM CLASSIFIER

$$P(f_1, f_2, ...., f_n | c) = P(f_1|c) \cdot P(f_2|c) \cdot ... \cdot P(f_n|c)$$

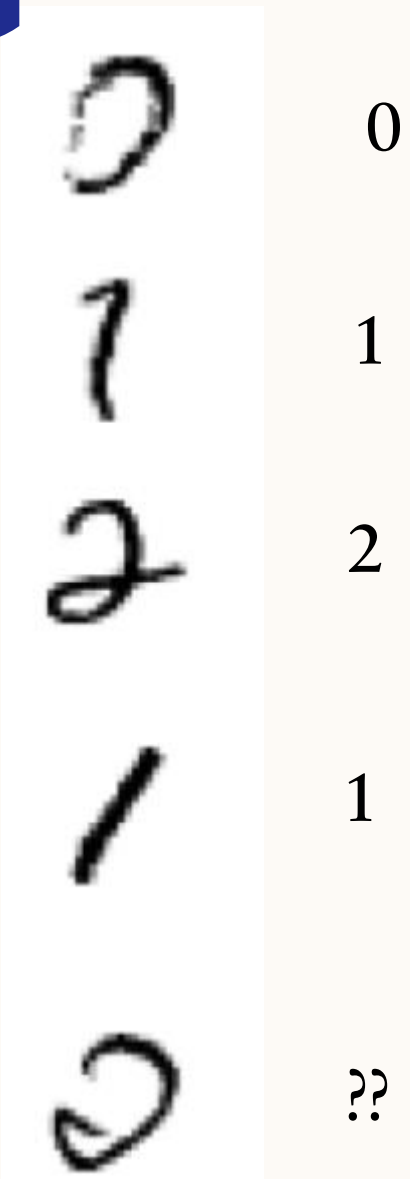
- What are our classes (what are we trying to classify)?
  - Spam, not spam
- What are our features (attributes used to classify)?
  - Certain words: free!, top secret
  - Text patterns: CAPS, $ii
  - Non-text: sender_in_contacts

| |
|---|
| Dear Sir.<br>First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. … |
| TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.<br>99 MILLION EMAIL ADDRESSES FOR ONLY $99 |
| Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power |

# NAÏVE BAYES HANDWRITING DIGIT RECOGNITION

$$P(f_1, f_2, ...., f_n | c) = P(f_1|c) \cdot P(f_2|c) \cdot ... \cdot P(f_n|c)$$

- What are our classes?
  - Digits 0-9
- What are our features?
  - Pixels
  - Shape patterns: number of components, aspect ratio, number of loops



# SELECTING A CLASS

$$P(f_1, f_2, ...., f_n|c) = P(f_1|c) \cdot P(f_2|c) \cdot ... \cdot P(f_n|c)$$

For a chosen class *c*:

$$c_{NB} = \underset{c \in C}{argmax}\ P(c) \prod_{f \in F} P(f|c)$$

# SETTING UP A NAÏVE BAYES CLASSIFIER

$$c_{NB} = \underset{c \in C}{argmax}\ P(c) \prod_{f \in F} P(f|c)$$

Certain words: free!, top secret
Text patterns: CAPS, $ii
Non-text: sender_in_contacts

spam

$P(spam)P(\neg\text{"free!"}|spam)P(\text{"top secret"}|spam)\ldots$ .73

$P(not\_spam)P(\neg\text{"free!"}|not\_spam)P(\text{"top secret"}|not\_spam)\ldots$ .14

| |
|---|
| Dear Sir.<br>First, I must solicit your confidence in this transaction, this is by virture of its nature as being utterly confidencial and top secret. … |
| TO BE REMOVED FROM FUTURE MAILINGS, SIMPLY REPLY TO THIS MESSAGE AND PUT "REMOVE" IN THE SUBJECT.<br>99 MILLION EMAIL ADDRESSES FOR ONLY $99 |
| Ok, Iknow this is blatantly OT but I'm beginning to go insane. Had an old Dell Dimension XPS sitting in the corner and decided to put it to use, I know it was working pre being stuck in the corner, but when I plugged it in, hit the power |

# SETTING UP A NAÏVE BAYES CLASSIFIER

- How do we learn a naïve bayes classifier?
  - $P(C|f_1 \dots fn) = P(C) \prod_{f \in F} P(f|C)$
  - Where each $f$ is 0/1 depending on whether it is present
  - Compute posterior distribution over class Y

3) Normalize by dividing step 1 by step 2

$$P(Y, f_1 \dots f_n) = \begin{bmatrix} P(y_1, f_1 \dots f_n) \\ P(y_2, f_1 \dots f_n) \\ \vdots \\ P(y_k, f_1 \dots f_n) \end{bmatrix} \Rightarrow \frac{\begin{bmatrix} P(y_1) \prod_i P(f_i|y_1) \\ P(y_2) \prod_i P(f_i|y_2) \\ \vdots \\ P(y_k) \prod_i P(f_i|y_k) \end{bmatrix}}{P(f_1 \dots f_n)}$$

1) Get joint probability of the class and evidence for each class

2) Sum them to get the probability of the evidence