

Welcome!

# A Brief Peek Into All Things Bayes

# What's up, Doc?

- What is Bayes' theorem?
- How does Bayesian classification work?
- When should Bayesian classifiers be used?
- Let's get our hands dirty!

Will you be my  
Bayes?

Trump versus Clinton

Trump  
 $P(T)$

Clinton  
 $P(C)$

# The Issue

Erecting a wall on the entire US-Mexico border

$$P(W)$$



$$P(W')$$

$$P(W \mid T) = 0.75$$

$$P(W' \mid T) = 0.25$$

$$P(W \mid C) = 0.05$$

$$P(W' \mid C) = 0.95$$

What's the probability a wall  
will be built if we don't know  
who becomes president?

$$\begin{aligned} P(W) = & \\ & P(C) \times P(W \mid C) \\ & + \\ & P(T) \times P(W \mid T) \end{aligned}$$

$$P(W) = \\ 0.5 \times 0.05 \\ + \\ 0.5 \times 0.75$$



$P(W) =$

0.025

+

0.375

$$P(W) = 0.4$$

Who's the President?

$P(T | W)$  or  $P(C | W)$

$$P(T | W) = \frac{P(W | T) \times P(T)}{P(W)}$$

$$P(T \mid W) = \frac{0.75 \times 0.5}{0.4}$$

$$P(T \mid W) = 0.9375$$

$$P(C \mid W) = \frac{P(C \mid T) \times P(C)}{P(W)}$$



$$P(C \mid W) = \frac{0.05 \times 0.5}{0.4}$$

$$P(C \mid W) = 0.0625$$

Trump

But why?

$$P(A \cap B) = P(B) \times P(A | B)$$

Basic Axiom of Probability

$$P(W \cap T) = P(T) \times P(W | T)$$

$$P(W \cap T) = P(W) \times P(T | W)$$

$$P(W) \times P(T | W)$$

$$=$$

$$P(T) \times P(W | T)$$



$$P(T | W)$$

$$=$$

$$P(T) \times P(W | T) / P(W)$$

$$P(A \mid B)$$

$$=$$

$$P(A) \times P(B \mid A) / P(B)$$

How does this relate  
to machine learning?

We have 3 features: age,  
college GPA, and  
gender.

What is the probability that the person associated with these features is a lawyer versus a journalist?

$$\begin{aligned} &P(\text{lawyer} \mid \text{age, gpa, gender}) \\ &= \\ &\frac{p(\text{age, gpa, gender} \mid \text{lawyer}) \times p(\text{lawyer})}{p(\text{age, gpa, gender})} \end{aligned}$$

$$\begin{aligned} &P(\text{journalist} \mid \text{age, gpa, gender}) \\ &= \\ &p(\text{age, gpa, gender} \mid \text{journalist}) \times p(\text{journalist}) \\ &\quad / p(\text{age, gpa, gender}) \end{aligned}$$

$p(\text{age, gpa, gender} \mid \text{journalist})$   
or  
 $p(\text{age, gpa, gender} \mid \text{lawyer})$



$p(\text{age} \mid \text{journalist})$

$p(\text{gpa} \mid \text{journalist})$

$p(\text{gender} \mid \text{journalist})$

# Calculating Conditional Probabilities

instantiate number of features

instantiate the probability of being a journalist to 0.5

instantiate two dictionaries to calculate conditional probabilities for each feature

for each feature set in the dataset

    for each feature

        increment the count of that feature for its class

    increment count of features encountered in class

divide the count of each individual feature by the total

# Types of Classifiers in sklearn

Bernoulli

# Multinomial

# When should I use Bayesian classification?

- When you are working with small amounts of data.
- When you are working with nominal values.
- When it's OK to make the assumption that each of the features are independent of each other.
- When you'd like to be able to explain to people what is happening under the hood with the technique

# Extra Resources

- Harrington, P. (2012). Machine Learning in Action. Shelter Island, NY: Manning Publications.
- Academy, Khan. (2013). Conditional Probability Explained Visually (Bayes Theorem Formula). Retrieved from <https://www.youtube.com/watch?v=Zxm4Xxvzohk>
- Monk, Mathematical. (2011). Naive Bayes Classification. Retrieved from <https://www.youtube.com/watch?v=8yvBqhm92xA>
- Monk, Mathematical. (2011). Naive Bayes Classification. Retrieved from <https://youtube.com/qX7n53NWYl4>
- [Jupyter Notebook for this Presentation](#)

[safiarocks.com/talks](https://safiarocks.com/talks)



git.io/vwZWj