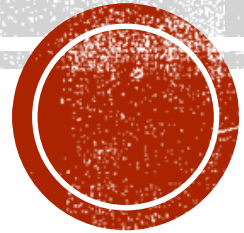
NAÏVE BAYES



INTRODUCTION

- Naïve Bayes is a subset of machine learning

WHAT IS MACHINE LEARNING?

- Gives "computers the ability to learn without being explicitly programmed"
  - Term coined by Arthur Samuel in 1959
  - Arthur Samuel taught a computer to play checkers and won against a renowned player in 1962
- Algorithms
  - A set of guidelines that tells a computer how to perform a task

TYPES OF TASKS IN MACHINE LEARNING

- Supervised learning
  - Algorithms that use labeled data with known possible outputs and training data contains "correct" answers
- Unsupervised learning
  - Unlabeled data where the algorithm has no background information to label the data and must determine a label based on features of the data points in the set
- Reinforcement learning
  - Algorithms that make decisions based on past decisions of the data

WHAT IS NAÏVE BAYES?

- Classification algorithm based on Bayes' theorem
- What is Bayes' theorem?
  - A way of finding probability when other probabilities are known
  - How often something happens given that something else happened

EXAMPLE OF BAYES' THEOREM

- 50% of rainy days start off cloudy

  P(Cloud|Rain) = 50%

- 40% of mornings are cloudy

  P(Cloud) = 40%

- For a particular season, 10% of days are rainy each month

  P(Rain) = 10%

$$P(Rain|Cloud) = \frac{P(Rain) * P(Cloud|Rain)}{P(Cloud)} \qquad P(Rain|Cloud) = \frac{0.1 * 0.5}{0.4} = .125$$

12.5% chance of rain

EXAMPLE OF NAÏVE BAYES

- 50% of the fruits are bananas
  P(banana) = 0.5
- 30% are oranges
  P(orange) = 0.3
- 20% are other fruits
  P(other) = 0.2

Fruit	Long	Sweet	Yellow	Total
Banana	400	350	450	500
Orange	0	150	300	300
Other	100	150	50	200
Total	**500**	**650**	**800**	**1000**

- P(long) = 0.5
- P(sweet) = 0.65
- P(yellow) = 0.8

- Out of 500 bananas, 400 (80%) are long, 350 (70%) are sweet, and 450 (90%) are yellow
- Out of 300 oranges, 0 (0%) are long, 150 (50%) are sweet, and 300 (100%) are yellow
- From the other 200 fruits, 100 (50%) are long, 150 (75%) are sweet, and 50 (25%) are yellow

Predict what a long, sweet, yellow fruit is…

NAÏVE BAYES' EXAMPLE (CONT'D)

$$P(\text{banana}|\text{long,sweet,yellow}) = \frac{P(\text{banana}) * P(\text{long}|\text{banana}) * P(\text{sweet}|\text{banana}) * P(\text{yellow}|\text{banana})}{P(\text{long}) * P(\text{sweet}) * P(\text{yellow})}$$

$$\frac{0.5*0.8*0.7*0.9}{0.5 * 0.65 * 0.8} = 0.969$$

$$P(\text{orange}|\text{long,sweet,yellow}) = 0$$

$$P(\text{other}|\text{long,sweet,yellow}) = \frac{P(\text{other}) * P(\text{long}|\text{other}) * P(\text{sweet}|\text{other}) * P(\text{yellow}|\text{other})}{P(\text{long}) * P(\text{sweet}) * P(\text{yellow})}$$

$$\frac{0.2*0.5*0.75*0.25}{0.5 * 0.65 * 0.8} = 0.072$$

PROS AND CONS OF NAÏVE BAYES

- Pros
  - Simple to build
  - Easy to train
- Cons
  - Assumes that all features are independent

- False Positives and False Negatives