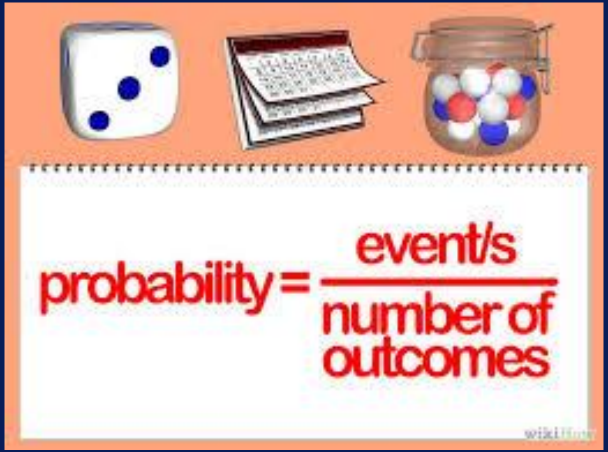
Naïve Bayes

After this video you will be able to..

- Discuss how a Naïve Bayes model works for classification
- Define the components of Bayes' Rule
- Explain what the 'naïve' means in Naïve Bayes

Naïve Bayes Overview



- Probabilistic approach to classification
  - Relationships between input features and class expressed as probabilities
  - Label for sample is class with highest probability given input

Naïve Bayes Classifier

Classification Using Probability

+

Bayes Theorem

+

Feature Independence Assumption

Probability of Event

Probability is measure of how likely an event is

Probability of Event 'A' Occurring

$$P(A) = \frac{\text{\# ways for A}}{\text{\# possible outcomes}}$$

Probability of Event

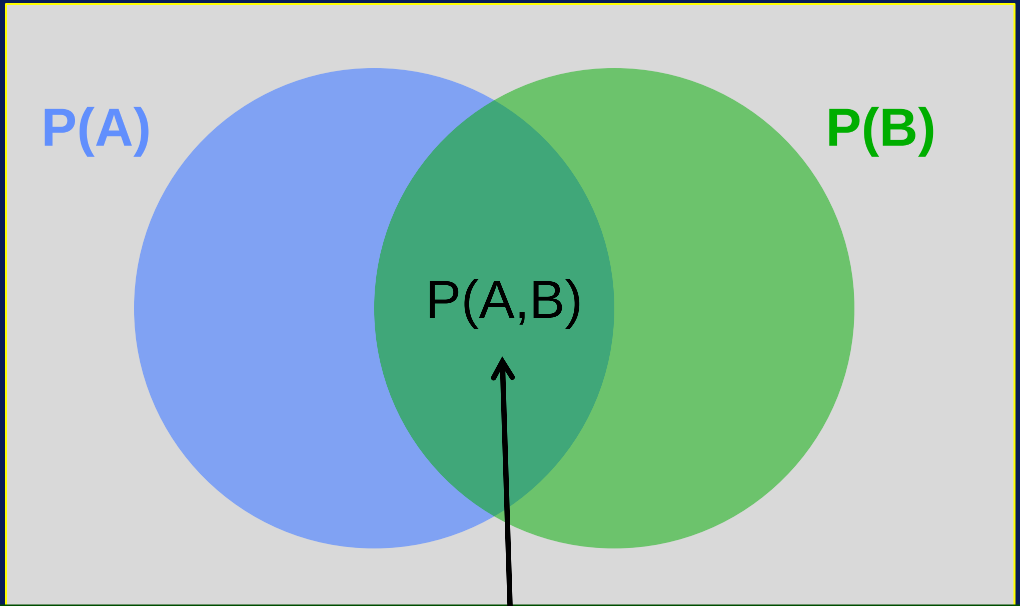
What is the probability of rolling a die and getting 6?

Probability of Rolling 6 on a Die

$$P(6) = \frac{\text{\# ways for getting 6}}{\text{\# possible outcomes}} = \frac{1}{6}$$

Joint Probability

Probability of events A and B occurring together



Joint Probability of A and B

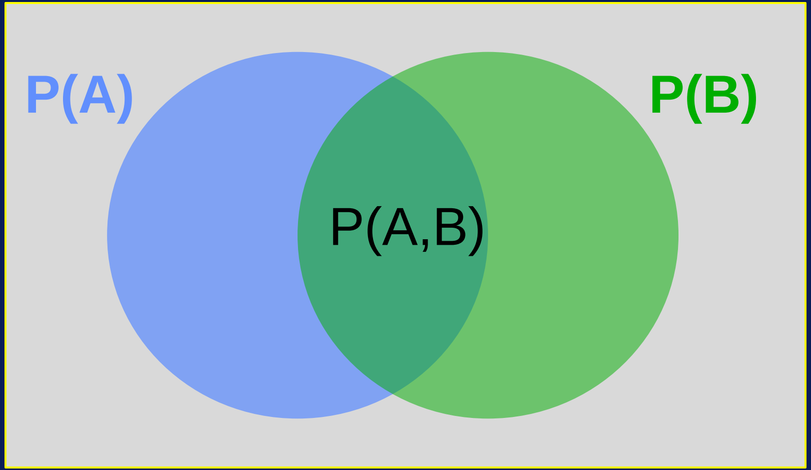
Joint Probability Example

What is the probability of two 6's when rolling two dice?

Probability of Rolling Two 6's

$$P(A,B) = P(A) * P(B) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$$

Conditional Probability



Probability of event A occurring, given that event B occurred

$$P(A \mid B) = \frac{P(A,B)}{P(B)}$$

Conditional Probability

Bayes' Theorem

- Relationship between P(B | A) and P(A | B) can be expressed through Bayes' Theorem

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

Bayes' Theorem

Classification with Probabilities

Given features X={X1,X2,…,Xn}, predict class C

Do this by finding value of C that maximizes P(C | X)

$$\left.\begin{array}{l} P(C_1|X) \\ P(C_2|X) \\ \quad\vdots \\ P(C_k|X) \end{array}\right\} \max \Rightarrow \text{Class Label}$$

Bayes' Theorem for Classification

- **But estimating P(C|X) is difficult**
- **Bayes' Theorem to the rescue!**
  - Simplifies problem

Bayes' Theorem for Classification

Posterior Probability

Class-Conditional Probability

Prior Probability

$$P(C \mid X) = \frac{P(X \mid C) * P(C)}{P(X)}$$

Probability of observing values for input features

Bayes' Theorem for Classification

Need to calculate this

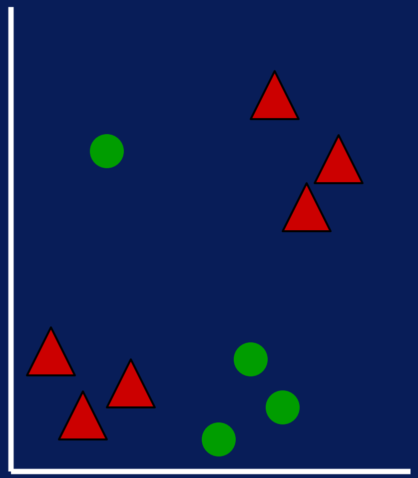
Can be estimated from data!

$$P(C \mid X) = \frac{P(X \mid C) * P(C)}{P(X)}$$

Constant (can be ignored)

To get P(C | X), only need to find P(X | C) and P(C), which can be estimated from the data!

Estimating P(C)



P(●) = 4/10 = 0.4

P(▲) = 6/10 = 0.6

To estimate P(C), calculate fraction of samples for class C in training data.

Estimating P(X | C)

Independence Assumption

- Features are independent of one another:

$$P(X_1,X_2,\ldots,X_n \mid C) = P(X_1|C) * P(X_2 \mid C) * \ldots * P(X_n \mid C)$$

To estimate P(X | C), only need to estimate $P(X_i \mid C)$ individually ➔ Much simpler!

Estimating $P(X_i \mid C)$

Home Owner	Marital Status	Loan Default
Yes	Single	No
No	Married	No
No	Single	No
Yes	Married	No
No	Divorced	Yes
No	Married	No
Yes	Divorced	No
No	Single	Yes
No	Married	No
No	Single	Yes

P(Home Owner = Yes | No) = 3/7 = 0.43

P(Marital Status = Single| Yes) = 2/3 = 0.67

Source: http://www-users.cs.umn.edu/~kumar/dmbook/index.php

Naïve Bayes Classification

- **Fast and simple**
- **Scales well**
- **Independence assumption may not hold true**
  - In practice, still works quite well
- **Does not model interactions between features**

Naïve Bayes Classifier

Classification Using Probability

+

Bayes Theorem

+

Feature Independence Assumption