Bayes Classifier

Will it rain?

- You and your friends are planning a picnic, but it is cloudy in the morning. You know that 50% of rainy days are cloudy in the morning. Cloudy mornings are common though, 40% of the days start off cloudy. Also, only 10% of days have rain.
- $\Pr(\text{Cloud} \mid \text{Rain}) = 0.5$
- $\Pr(\text{Cloud}) = 0.4$
- $\Pr(\text{Rain}) = 0.1$
- $\Pr(\text{Cloud} \mid \text{Rain}) = ?$
- $\Pr(\text{Rain} \mid \text{Cloud}) = \frac{\Pr(\text{Cloud}|\text{Rain}) \cdot \Pr(\text{Rain})}{\Pr(\text{Cloud})} = \frac{0.5 \cdot 0.1}{0.4} = 0.125$

Terminology

- Given a hypothesis h and data D:

$$\Pr(h \mid D) = \frac{\Pr(D \mid h) \cdot \Pr(h)}{\Pr(D)}$$

- $\Pr(h)$: independent probability of h: prior probability
- $\Pr(D)$: independent probability of D
- $\Pr(D \mid h)$: conditional probability of D given h: likelihood
- $\Pr(h \mid D)$: conditional probability of h given D: posterior probability

Stiff Neck

- Given:
  - A doctor knows that meningitis causes stiff necks 50% of the time
  - Prior probability of any patient having meningitis is 1/50,000
  - Prior probability of any patient having stiff neck is 1/20
- If a patient has a stiff neck, what is the probability that they have meningitis?
- $\Pr(M \mid S) = ?$

$$\Pr(M \mid S) = \frac{\Pr(S \mid M) \cdot \Pr(M)}{\Pr(S)} = \frac{0.5 \cdot 1/50000}{1/20} = 0.0002$$

Exercise

- Suppose the fraction of undergraduate students who smoke is 15% and the fraction of graduate students who smoke is 23%. If 1/5 of the university students are graduate students and the rest are undergraduates, what is the probability that a student who smokes is a graduate student?

- We know:
  - $\Pr(S \mid UG) = 0.15$
  - $\Pr(S \mid G) = 0.23$
  - $\Pr(G) = 0.2$
  - $\Pr(UG) = 0.8$
- Note: $\Pr(S) = \Pr(S \mid UG) \cdot \Pr(UG) + \Pr(S \mid G) \cdot \Pr(G)$
- We want to know $\Pr(G \mid S)$

$$\Pr(G \mid S) = \frac{\Pr(S \mid G) \cdot \Pr(G)}{\Pr(S)} = \frac{0.23 \cdot 0.2}{0.15 \cdot 0.8 + 0.23 \cdot 0.2} = 0.277$$

Who will win?

- Consider a football game between rival teams: Team 0 and Team 1. Suppose Team 0 wins 65% of the time and Team 1 wins the remaining matches. Among the games won by Team 0, only 30% of them come from playing on Team 1's football field.

  On the other hand, 75% of the victories for Team 1 are obtained while playing at home. If Team 1 is to host the next match between the two teams, which team is most likely to emerge as the winner?
- Winner $= Y$
- Host $= X$
- We know:
  - $\Pr(Y = 0) = 0.65$
  - $\Pr(Y = 1) = 1 - \Pr(Y = 0) = 0.35$
  - $\Pr(X = 1 \mid Y = 1) = 0.75$
  - $\Pr(X = 1 \mid Y = 0) = 0.3$
- We want $P(Y = 1 \mid X = 1)$

$$\begin{aligned}
\Pr(Y = 1 \mid X = 1) &= \frac{\Pr(X = 1 \mid Y = 1) \cdot \Pr(Y = 1)}{\Pr(X = 1)} \\
&= \frac{\Pr(X = 1 \mid Y = 1) \cdot \Pr(Y = 1)}{\Pr(X = 1 \mid Y = 1) \cdot \Pr(Y = 1) + \Pr(X = 1 \mid Y = 0) \cdot \Pr(Y = 0)} \\
&= \frac{0.75 \cdot 0.35}{0.75 \cdot 0.35 + 0.3 \cdot 0.65} \\
&= 0.5738
\end{aligned}$$

Bayesian Classifiers

- Consider each attribute and class label as random variables
- Given a test record with attributes $(A_1, A_2, \ldots, A_n)$

  - Goal to predict class C
  - Specifically, we need to find the value of C that maximizes $\Pr(C \mid A_1, A_2, \ldots, A_n)$
- Can we estimate $\Pr(C \mid A_1, A_2, \ldots, A_n)$ directly from training data?
- Training: Learn posterior probabilities $\Pr(Y \mid X)$
- Testing Find class Y that maximizes $P(Y \mid X')$ where X' is the sample / record
- Suppose we are given a test record with the following:

$$X = \{\text{Body Temperature} = \text{Cold-Blooded}, \text{Skin Cover} = \text{Scales}, \text{Has Legs} = \text{Yes}\}$$

- To classify X with one of the labels (Reptile / Not), we need to calculate the posterior probabilities $\Pr(\text{Reptile} \mid X)$ and $\Pr(\text{Not} \mid X)$ based on the information in the training data
- Estimating the posterior probabilities accurately for every possible combination of class label and attribute value is a difficult problem because it requires a very large training set, even for a moderate number of attributes
- The Bayes Theorem is useful because it allows us to express the posterior probability in terms of the prior probability $\Pr(Y)$, the class-conditional probability $\Pr(X \mid Y)$, and the evidence $P(X)$

Naïve Bayes Classifiers

- A naïve Bayes classifier estimates the class-conditional probability by assuming that the attributes are conditionally independent, given the class label y.
- The "naive" component refers to the assumption that the presence of features does not affect the presence of other features

$$P(Y \mid X) = \frac{\Pr(X \mid Y) \cdot \Pr(Y)}{\Pr(X)}$$

- Attribute set $X = \{X_1, X_2, \ldots, X_d\}$
- $\Pr(X \mid Y)$ can be very difficult to calculate
- What are the odds that we have seen all features with the same values as X, from training?
- To address this concern, we need to discuss conditional independence