

Bayes Classifier

Conditional Independence

- A And B are conditionally independent of C if $\Pr(A, B | C) = \Pr(A | C) \cdot \Pr(B | C)$
- Another way to view this is $\Pr(A | B, C) = \Pr(A | C)$ and $\Pr(B | A, C) = \Pr(B | C)$
- That is, if I know C , then whether or not I know B does not influence my probability of A
- Assume that we have two coins to choose from
 - Coin 1: Fair coin
 - Coin 2 : Unfair coin (heads on both sides)
- I randomly reach into a bag and pull out a coin and flip it twice. I have the following events to consider:
 - A : First coin toss results in a head
 - B : Second coin toss results in a head
 - C : I selected coin 1 to flip
- If I know that I selected coin 1 (C), then knowing the outcome A does not influence my probability of B , $\Pr(B | A, C) = \Pr(B | C)$

Independence

- Conditionally independent does not imply independent
- Using the previous example with coins, consider the following probabilities

$$\Pr(A) = \Pr(A | C) \cdot \Pr(C) + \Pr(A | C^C) \cdot \Pr(C^C) = \frac{3}{4}$$

$$\Pr(B) = \Pr(B | C) \cdot \Pr(C) + \Pr(B | C^C) \cdot \Pr(C^C) = \frac{3}{4}$$

$$\begin{aligned} \Pr(A, B) &= \Pr(A, B | C) \cdot \Pr(C) + \Pr(A, B | C^C) \cdot \Pr(C^C) \\ &= \Pr(A | C) \cdot \Pr(B | C) \cdot \Pr(C) + \Pr(A | C^C) \cdot \Pr(B | C^C) \cdot \Pr(C^C) \\ &= \frac{5}{8} \end{aligned}$$

- Here, $\Pr(A, B) = \frac{5}{8} \neq \Pr(A) \cdot \Pr(B) = \frac{9}{16}$
- These are not independent, knowing that A occurred increases the chance of B occurring. However, given C , A and B are independent.

Naïve Bayes Classifier

- We want to calculate $\Pr(Y | X)$ with:

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

- Assuming conditional independence,

$$\Pr(X \mid Y) = \prod_{i=1}^d \Pr(X_i \mid Y)$$

- So, we can now calculate $\Pr(Y \mid X)$ as:

$$\Pr(Y \mid X) = \frac{\Pr(Y) \cdot \prod_{i=1}^d \Pr(X_i \mid Y)}{\Pr(X)}$$

Bayesian Classifiers

- Since the reason to calculate $\Pr(Y \mid X)$ is to classify an instance with the highest probability, we can perform one extra shortcut
- Assume we are given a test instance X with the feature vector $[1, 2, 3]$, we need to label it as class $C1$ or class $C2$
- So, we need to determine:

$$\hat{y} = \operatorname{argmax}_{y \in Y} (\Pr(Y = y \mid X))$$

- In other words, we need to compare

$$\Pr(Y = C1 \mid X) \quad \text{against} \quad \Pr(Y = C2 \mid X)$$

- That is,

$$\max \left(\frac{\Pr(Y = C1) \cdot \prod_{i=1}^d \Pr(X_i \mid Y = C1)}{\Pr(X)}, \frac{\Pr(Y = C2) \cdot \prod_{i=1}^d \Pr(X_i \mid Y = C2)}{\Pr(X)} \right)$$

- Notice that $\Pr(X)$ does not change between different classes. It is the same feature vector, so we can simplify to finding which of the following is larger:

$$\Pr(Y = C1) \cdot \prod_{i=1}^d \Pr(X_i \mid Y = C1) \quad \text{or} \quad \Pr(Y = C2) \cdot \prod_{i=1}^d \Pr(X_i \mid Y = C2)$$

- We can estimate $\Pr(X_i = x \mid Y = y)$ based on the fraction of training instances with class y and also has $X_i = x$
- Consider the following table:

ID	Home Owner	Martial Status	Owns a Cat
1	Yes	Single	No
2	No	Married	No
3	No	Single	No
4	Yes	Married	No
5	No	Divorced	Yes
6	No	Married	No
7	Yes	Divorced	No
8	No	Single	Yes
9	No	Married	No
10	No	Single	Yes

- How would we classify someone who is not a home owner, and is single?

$$\begin{aligned}\Pr(\text{Yes} \mid X) &= \frac{\Pr(\text{Yes}) \cdot \Pr(\text{HomeOwner} = \text{No} \mid \text{Yes}) \cdot \Pr(\text{Single} \mid \text{Yes})}{\Pr(X)} \\ &= \frac{\frac{3}{10} \cdot 1 \cdot \frac{2}{3}}{\Pr(X)} \\ &= \frac{0.2}{\Pr(X)}\end{aligned}$$

$$\begin{aligned}\Pr(\text{No} \mid X) &= \frac{\Pr(\text{No}) \cdot \Pr(\text{HomeOwner} = \text{No} \mid \text{No}) \cdot \Pr(\text{Single} \mid \text{No})}{\Pr(X)} \\ &= \frac{\frac{7}{10} \cdot \frac{4}{7} \cdot \frac{2}{7}}{\Pr(X)} \\ &\approx \frac{0.114}{\Pr(X)}\end{aligned}$$

- So, we would classify our instance as yes, that is, they would own a cat.

Missing Values

- Consider the following table:

TID	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Given a test record with the attributes:

$$X = \{\text{Refund} = \text{Yes}, \text{Status} = \text{Divorced}\}$$

- Now, think about finding $\Pr(X \mid \text{Yes})$

$$\begin{aligned}\Pr(X \mid Y) &= \Pr(\text{Refund} = \text{Yes} \mid \text{Yes}) \cdot \Pr(\text{Status} = \text{Divorced} \mid \text{Yes}) \\ &= 0 \cdot \frac{1}{3} \\ &= 0\end{aligned}$$

- Thus, the naïve Bayes classifier will not be able to classify the record

- Consider the new test record with:

$$X = \{\text{Refund} = \text{Yes}, \text{Status} = \text{Widowed}, \text{Income} = 120K\}$$

- If one of the conditional probabilities is zero, then the entire expression becomes zero, $\Pr(X | \text{Yes}) = 0$
- If the training examples do not cover many of the attribute values, the expression also becomes zero. $\{\text{Status} = \text{Widowed}\}$ is not covered, so $\Pr(X | \text{No}) = 0$
- How do we address these special cases?

Smoothing / Estimating

- To address the issues of 0, we can perform smoothing. We will use the m -estimate approach
- We can calculate $\Pr(x_i | y) = \frac{n_c + mp}{n + m}$
 - n : total number of instances from class y
 - n_c : number of training instances from class y that has value x_i
 - m : a parameter known as the equivalent sample size (often between 1 and 3)
 - p : a user-specified parameter, which can be seen as the prior of observing the attribute value x_i among instances with class y
 - * If we assume a uniform distribution for values in a given attribute, then $p = \frac{1}{t}$ where t is the number of unique values that the attribute can have
 - If there is no training set:

$$\Pr(x_i | y) = \frac{0 + mp}{0 + m} = p$$

- If there is a training set, but no instances with class y and value x_i :

$$\Pr(x_i | y) = \frac{0 + mp}{n + m}$$

- Using smoothing, and $m = 2$, calculate $\Pr(X | \text{Yes})$ for:

$$X = \{\text{Refund} = \text{Yes}, \text{Status} = \text{Divorced}\}$$

$$\begin{aligned} \Pr(X | Y) &= \Pr(\text{Refund} = \text{Yes} | \text{Yes}) \cdot \Pr(\text{Status} = \text{Divorced} | \text{Yes}) \\ &= \frac{0 + (2 \cdot \frac{1}{2})}{3 + 2} \cdot \frac{1 + (2 \cdot \frac{1}{3})}{3 + 2} \\ &= \frac{1}{5} \cdot \frac{1}{3} \\ &= \frac{1}{15} \end{aligned}$$