DSC 40A - Discussion 07 - Naive Bayes and Conditional Independence

March 10, 2020

Problem 1.

In parts of the world other than San Diego, the weather changes from day to day. In these places, people try to guess tomorrow's weather using the current conditions.

Weather data for 20 random days in Columbus, Ohio are recorded below, along with the next day's weather (rainy, cloudy, or sunny).

Suppose that today's humidity is $> 50\%$, the temperature is hot, and the air pressure is low. Use naïve Bayes to predict whether tomorrow will be rainy, cloudy, or sunny. Show your work.

Next Day's Weather	**Humidity**	**Temperature**	**Air Pressure**
Rainy	$> 50\%$	Cool	Low
Rainy	$> 50\%$	Hot	Low
Rainy	$> 50\%$	Cool	Low
Rainy	25%-50%	Hot	High
Rainy	25%-50%	Hot	Low
Rainy	25%-50%	Cool	Low
Rainy	25%-50%	Cool	Low
Rainy	$< 25\%$	Cool	Low
Rainy	$< 25\%$	Hot	Low
Rainy	$< 25\%$	Hot	High
Cloudy	$> 50\%$	Cool	Low
Cloudy	$> 50\%$	Cool	Low
Cloudy	25%-50%	Hot	High
Cloudy	$< 25\%$	Cool	High
Cloudy	$< 25\%$	Cool	Low
Sunny	$> 50\%$	Cool	Low
Sunny	$> 50\%$	Hot	High
Sunny	$> 50\%$	Cool	High
Sunny	25%-50%	Hot	High
Sunny	$< 25\%$	Hot	High

Prediction: tommorow will be ● rainy ○ cloudy ○ sunny

Solution: Naïve Bayes calls for computing three things and seeing which is the largest:

$$\begin{aligned}
P(\text{Rainy} \mid {>}50\% \text{ and hot and low}) &\propto P({>}50\% \mid \text{Rainy}) \cdot P(\text{hot} \mid \text{Rainy}) \cdot P(\text{low} \mid \text{Rainy}) \cdot P(\text{Rainy}) \\
&\approx \frac{3}{10} \cdot \frac{5}{10} \cdot \frac{8}{10} \cdot \frac{10}{20} \\
&= \frac{1200}{20000} \\
P(\text{Cloudy} \mid {>}50\% \text{ and hot and low}) &\propto P({>}50\% \mid \text{Cloudy}) \cdot P(\text{hot} \mid \text{Cloudy}) \cdot P(\text{low} \mid \text{Cloudy}) \cdot P(\text{Cloudy}) \\
&\approx \frac{2}{5} \cdot \frac{1}{5} \cdot \frac{3}{5} \cdot \frac{5}{20} \\
&= \frac{30}{2500} = \frac{240}{20000} \\
P(\text{Sunny} \mid {>}50\% \text{ and hot and low}) &\propto P({>}50\% \mid \text{Sunny}) \cdot P(\text{hot} \mid \text{Sunny}) \cdot P(\text{low} \mid \text{Sunny}) \cdot P(\text{Sunny}) \\
&\approx \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{5}{20} \\
&= \frac{45}{2500} = \frac{360}{20000}
\end{aligned}$$

So it is most likely to be rainy tomorrow.

Problem 2.

This problem will illustrate that independence and conditional independence are not related, in the sense that neither property implies the other.

Consider the sample space $S = \{1, 2, 3, 4, 5, 6\}$ with associated probability distribution $P(s) = \frac{1}{6}$ for each s in S. You can think of S as representing the possible outcomes of rolling a single die.

For each part below, define events A, B, C in this sample space that satisfy the given requirements. Demonstrate that the requirements are satisfied by computing the appropriate probabilities. Make sure to choose events that are neither impossible nor certain, that is, $0 < P(A), P(B), P(C) < 1$.

a) A and B are independent.
A and B are conditionally independent given C.

Solution:

$$A = \{1, 2, 3\}, B = \{1, 4\}, C = \{1, 2, 4, 5\}$$

We know that A and B are independent because

$$\begin{aligned}
P(A) &= \frac{1}{2} \\
P(B) &= \frac{1}{3} \\
P(A \cap B) = P(\{1\}) &= \frac{1}{6}
\end{aligned}$$

so $P(A \cap B) = P(A) * P(B) = \dfrac{1}{6}$.

We know that A and B are conditionally independent given C because

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{1,2\})}{P(\{1,2,4,5\})} = \frac{1}{2}$$
$$P(B|C) = \frac{P(B \cap C)}{P(C)} = \frac{P(\{1,4\})}{P(\{1,2,4,5\})} = \frac{1}{2}$$
$$P((A \cap B)|C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{P(\{1\})}{P(\{1,2,4,5\})} = \frac{1}{4}$$

so $P((A \cap B)|C) = P(A|C) * P(B|C) = \frac{1}{4}$.

Problem 3.

A box contains two coins: a regular coin and one fake two-headed coin ($P(H) = 1$). I choose a coin at random and toss it twice. Define the following events.

1. A= First coin toss results in a heads (H).
2. B= Second coin toss results in a heads (H).
3. C= Coin 1 (regular) has been selected.

Find $P(A|C), P(B|C), P(A \cap B|C), P(A), P(B), P(A \cap B)$. Show that A and B are NOT independent, but they are conditionally independent given C.

Solution: Firstly, we know that $P(A|C) = P(B|C) = 1/2$ since a fair coin has equal probability of heads and tails. Also, given that Coin 1 is picked, we have $P(A \cap B|C) = (1/2)(1/2) = 1/4$ (the probability of getting two heads in two tosses of a fair coin). Therefore, $P(A \cap B|C) = P(A|C)P(B|C)$. We showed that **A and B are conditionally independent given C**.
Now, we need to show $P(A \cap B) \neq P(A)P(B)$.

To find $P(A), P(B)$ and $P(A \cap B)$, we use the law of total probability:

$$\begin{aligned} P(A) &= P(A|C)P(C) + P(A|\neg C)P(\neg C) \\ &= (1/2) * (1/2) + 1 * (1/2) \\ P(A) &= 3/4 \end{aligned}$$

Similarly,

$$\begin{aligned} P(B) &= P(B|C)P(C) + P(B|\neg C)P(\neg C) \\ &= (1/2) * (1/2) + 1 * (1/2) \\ P(B) &= 3/4 \end{aligned}$$

Finally,

$$\begin{aligned} P(A \cap B) &= P(A \cap B|C)P(C) + P(A \cap B|\neg C)P(\neg C) \\ &= P(A|C)P(B|C)P(C) + P(A|\neg C)P(B|\neg C)P(\neg C) \\ &= (1/2) * (1/2) * (1/2) + (1) * (1) * (1/2) \\ P(A \cap B) &= 5/8 \end{aligned}$$

Since $P(A \cap B) = 5/8 \neq P(A)P(B) = 9/16$, A **and** B **are not independent**. We can also justify this intuitively. For example, if we know that A has occurred (i.e., the first coin toss has resulted in heads), we would think that there is a higher chance of having chosen Coin 2 than Coin 1. This in turn increases the conditional probability of B occurring (in other words, $P(B|A) = 0.83 > P(B) = 0.75$), which suggests that A and B are not independent. On the other hand, A and B are independent given C (if we know that Coin 1 was selected). Generally speaking, conditional independence neither implies (nor is it implied by) independence.