Naive Bayes Classifier:

Theory explanation:

Framework:

In our exercise, we want to determine if given a description of the weather today, we decide whether to go play outdoor.

In this case, the features correspond to the different parameters describing the weather (Outlook, Temperature, Humidity, Windy). These parameters can take 2 or 3 different values. The target is the answer to the question “should we play outdoor?”, it can take two values (YES or NOT).

The objective of a classifier is to determine which class we should choose given a set of feature. For example, in our case, the goal would be to determine if we should play outdoor given a weather description (for example: outlook=Sunny, Temperature=hot, Humidity=High, Windy=TRUE).

How to determine the appropriate class?

What we actually want is: “find which class is the most appropriate for these feature”, or in other terms “which class has the highest probability given these feature”. In a math point of view, this correspond to finding which has the high probability between:

$P(\omega = YES \mid X)$ and $P(\omega = NO \mid X)$ where

- ω is the target and
- $X = (outlook = Sunny,\ Temperature = hot,\ Humidity = high,\ Windy = TRUE)$ is the vector of features describing the weather.

Comment: for convenience, we will write $X = (outlook = Sunny,\ Temperature = hot,$ $Humidity = high,\ Windy = TRUE)$ as $X = (X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)$

To do so, we calculate both of the probabilities.

How to calculate $P(\omega = YES \mid X)$?

Thanks to the Bayes theorem (make a link to an annex reminding the theorem) we can write the following equality:

$$P(\omega = YES \mid (X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)) = \frac{P(\omega = YES) \times P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4 \mid \omega = YES)}{P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4)}$$

Or:

$$P(\omega = YES \mid X) = \frac{P(\omega = YES) \times P(X \mid \omega = YES)}{P(X)}$$

Thanks to the Conditional probability rules, we can write $P(X \mid \omega = YES)$ as followed:

$$P(X \mid \omega = YES) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3, X_4 = x_4 \mid \omega = YES)$$

$$P(X|\omega = YES) = P(X_1 = x_1|\omega = YES) \times P(X_2 = x_2|\omega = YES, X_1 = x_1) \\ \times P(X_3 = x_3|\omega = YES, X_1 = x_1, X_2 = x_2) \\ \times P(X_4 = x_4|\omega = YES, X_1 = x_1, X_2 = x_2, X_3 = x_3)$$

But the Naïve Bayes Classifier use the Naïve assumption: "The effect of the value of a predictor x on a given class c is independent of the values of the other predictors". In other terms, a naïve bayes classifier assumes that the presence of a particular feature in a class in unrelated to the presence of any other feature. The assumption, also called **class conditional independence** can be expressed with a math formula:

$$P(x_i|\, \omega, x_j) = P(x_i|\, \omega)$$

Or:

$$P(X_i = x_i|\, \omega, X_j = x_j) = P(X_i = x_i|\, \omega) \ , \forall\, i \neq j$$

So now, we can simplify $P(X|\omega = YES)$

$$P(X|\omega = YES) = P(X_1 = x_1|\omega = YES) \times P(X_2 = x_2|\omega = YES) \times P(X_3 = x_3|\omega = YES) \\ \times P(X_4 = x_4|\omega = YES)$$

$$P(X|\omega = YES) = \prod_{i=1}^{4} P(X_i = x_i|\omega = YES)$$

Now we have:

$$P(\omega = YES\ |X\ ) = P(\omega = YES) \times \prod_{i=1}^{4} P(X_i = x_i|\omega = YES) \times \frac{1}{P(X)}$$

We proceed with the same method to get

$$P(\omega = NO\ |X\ ) = P(\omega = NO) \times \prod_{i=1}^{4} P(X_i = x_i|\omega = NO) \times \frac{1}{P(X)}$$

Now, we just need to calculate the two probabilities, and choose the highest one.

<u>Application</u>:

When we apply the classifier in real life, we actually do not need to calculate $\frac{1}{P(X)}$.

Actually, the Bayes classifier is the function that assigns a class label $\hat{y} = C_k$ for some k as follows:

$$\hat{y} = \underset{k \in \{1,\dots,K\}}{argmax}\, p(\omega_k) \times \prod_{i=1}^{n} P(x_i|\omega_k)$$

With

- $K : number\ of\ possible\ target$
- $\omega_k : the\ possible\ target$
- $n : number\ of\ features$

- x_i : $the\ possible\ feature$

Which is also equal to :

$$\hat{y} = \underset{k \in \{1,\dots,K\}}{argmax}\ p(\omega_k) \times \frac{1}{P(X)} \times \prod_{i=1}^{n} P(x_i|\omega_k) = \underset{k \in \{1,\dots,K\}}{argmax}\ P(\omega_k|X\ )$$

With :

- X : $vector\ of\ features$

Implementation:

Explanation of data structure , algo & program