Bayes Classifiers

Daniel Otero Fadul

Science Department
School of Engineering and Science

Artificial Intelligence

AI stands for "Artificial Intelligence." It refers to the ability of machines or computer programs to perform tasks that typically require human intelligence. These tasks include learning, perception, reasoning, problem solving, and understanding natural language.

Artificial Intelligence

Another definition of AI considers this field to be the study of agents that receive perceptions from the environment and perform actions. An agent is any system that perceives its environment through sensors and acts on it through actuators.

Artificial Intelligence

In general, AI is a vast field in which we can find a wide variety of techniques that can be used to give an agent a set of skills that enable it to perform a task that is normally performed by a human.

Naive Bayes Classifiers

Naive Bayes classifiers are a family of classifiers based on Bayes' theorem for assigning classes $\{C_k\}_{k=1}^{K}$ to a set of feature vectors x:

$$(x_1, x_2, \ldots, x_n) \to C_k.$$

These classifiers are considered "naive" because they assume that the features of vector x are conditionally independent given a class C_k.

Naive Bayes Classifiers

There are three types of classifiers:

- **Bernoulli:** Each component of vector x is a binary variable.
- **Multinomial:** Each component of x can take a certain set of values.
- **Gaussian:** It is assumed that each component of x is a continuous variable whose conditional distribution given a class C_k is a normal distribution.

Conditional Probability

Let A and B be two events belonging to a sample space Ω. The probability of event A occurring given that event B is known to have occurred is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

If A and B are independent, we have that $P(A|B) = P(A)$, which implies $P(A \cap B) = P(A)P(B)$.

Total Probability

Let $\{A_i\}_{i=1}^n$ be a collection of mutually exclusive events that form a partition of the sample space: $A_i \cap A_j = \emptyset$, $\forall i \neq j$ y $\cup_{i=1}^n A_i = \Omega$. Then, for any event $B \subset \Omega$, we have that

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n). \end{aligned}$$

Note that $\forall i, \ P(A_i) > 0$.

Bayes' Theorem

Let $\{A_i\}_{i=1}^{n}$ be a collection of mutually exclusive events that form a partition of the sample space: $A_i \cap A_j = \emptyset$, $\forall i \neq j$, $\cup_{i=1}^{n} A_i = \Omega$ y $P(A_i) > 0 \; \forall i$. Then, for any event $B \subset \Omega$such that $P(B) > 0$, we have that

$$\begin{aligned} P(A_i|B) &= \frac{P(B|A_i)P(A_i)}{P(B)} \\ &= \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n)}. \end{aligned}$$

Bayes' Theorem

When the partition of Ω consists of only two elements ($A \cup \bar{A} = \Omega$), we have the following special case:

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|\bar{A})P(\bar{A})}. \end{aligned}$$

Bayes' Theorem

Santa Claus packs the toys he delivers at Christmas in boxes containing twenty toys. Suppose that 60% of all the boxes Santa Claus delivers do not contain defective toys, 30% contain only one defective toy, and 10% contain two defective toys. A box is selected at random and two toys are chosen at random. The selected toys are not defective.

- What is the probability that there are no defective toys in the selected box?
- What is the probability that there is only one defective toy in the selected box?
- What is the probability that there are two defective toys in the selected box?

Bayes' Theorem

Let C_0, C_1 y C_2 be the events of box with zero defective toys, one defective toy, and two defective toys, respectively. Let J_0 be the event of choosing two non-defective toys from a box without replacement. Then, to answer the first question, we need to calculate $P(C_0|J_0)$:

$$\begin{aligned}
P(C_0|J_0) &= \frac{P(J_0|C_0)P(C_0)}{P(J_0|C_0)P(C_0)+P(J_0|C_1)P(C_1)+P(J_0|C_2)P(C_2)} \\
&= \frac{(1)\left(\frac{6}{10}\right)}{(1)\left(\frac{6}{10}\right)+\left(\frac{19}{20}\frac{18}{19}\right)\left(\frac{3}{10}\right)+\left(\frac{18}{20}\frac{17}{19}\right)\left(\frac{1}{10}\right)} \\
&= 0.631229\ldots
\end{aligned}$$

Naive Bayes Classifier

Let $x = (x_1, x_2, \ldots, x_n)$ be a vectos of features and let $\{C_k\}_{k=1}^{K}$ be a set of classes. When we use the Bayesian classifier as our model, we need to calculate the following probability:

$$\begin{aligned} P(C_k|x) &= \frac{P(x|C_k)P(C_k)}{P(x)} \\ &= \frac{P(x_1 \cap x_2 \cap \cdots \cap x_n \cap C_k)}{P(x)}. \end{aligned}$$

Naive Bayes Classifier

By the rule of multiplication, we have that the numerator is equal to

$$P(x_1 \cap x_2 \cap \cdots \cap x_n \cap C_k) = P(C_k)P(x_1|C_k)\cdots P(x_n|x_1 \cap x_2 \cap \cdots \cap C_k).$$

If the features are mutually and conditionally independent given a class C_k, then

$$P\left(x_i|x_1 \cap x_2 \cap \cdots \cap x_{i-1} \cap x_{i+1} \cap \cdots \cap x_n \cap C_k\right) = P(x_i|C_k).$$

Therefore, the first expression can be rewritten as follows:

$$P(x_1 \cap x_2 \cap \cdots \cap x_n \cap C_k) = P(C_k)P(x_1|C_k)\cdots P(x_n|C_k).$$

This condition does not normally occur in real life, which is why the Bayesian classifier is considered "naive." However, this method can perform quite well in a variety of situations.

Naive Bayes Classifier

To determine which class a vector x belongs to, the following calculation is performed:

$$\hat{y} = \underset{k=1,2,\ldots,K}{\operatorname{argmax}} P(C_k)P(x_1|C_k)\cdots P(x_n|C_k).$$