Machine Learning

Nirdosh Bhatnagar

Naïve Bayes' Classifier - Example

Example An elementary naïve Bayesian classifier is developed in this example. A data source outputs data with two attributes A_1 and A_2, and the data is divided into two classes ω_1 and ω_2. Attribute A_1 takes values from the set $\{a, b, c\}$, and the attribute A_2 takes values from the set $\{v, y, z\}$. The training data set is specified in the Table. The goal of the training data set is to learn a naïve Bayesian classifier.

Data Point	A_1	A_2	Class ω
1	a	y	ω_1
2	a	z	ω_1
3	c	y	ω_1
4	b	z	ω_1
5	b	v	ω_1
6	a	y	ω_2
7	c	v	ω_2
8	b	y	ω_2
9	c	z	ω_2
10	c	v	ω_2

Table. Training data set.

The following probabilities are readily computed from the table.

$$P(\omega = \omega_1) = 1/2, \quad \text{and} \quad P(\omega = \omega_2) = 1/2$$

$$\begin{aligned} P(A_1 = a \mid \omega = \omega_1) &= 2/5, \\ P(A_1 = b \mid \omega = \omega_1) &= 2/5, \\ P(A_1 = c \mid \omega = \omega_1) &= 1/5 \end{aligned}$$

$$\begin{aligned} P(A_1 = a \mid \omega = \omega_2) &= 1/5, \\ P(A_1 = b \mid \omega = \omega_2) &= 1/5, \\ P(A_1 = c \mid \omega = \omega_2) &= 3/5 \end{aligned}$$

$$\begin{aligned} P(A_2 = v \mid \omega = \omega_1) &= 1/5, \\ P(A_2 = y \mid \omega = \omega_1) &= 2/5, \\ P(A_2 = z \mid \omega = \omega_1) &= 2/5 \end{aligned}$$

$$\begin{aligned} P(A_2 = v \mid \omega = \omega_2) &= 2/5, \\ P(A_2 = y \mid \omega = \omega_2) &= 2/5, \\ P(A_2 = z \mid \omega = \omega_2) &= 1/5 \end{aligned}$$

The test sample is $x = (a_1, a_2) = (b, z)$. The goal is to determine the class ω of this data point. In this data point, $A_1 = b$, and $A_2 = z$. For class $\omega = \omega_1$ we compute

$$\begin{aligned} & P(\omega_1) \prod_{i=1}^{2} P(A_i = a_i \mid \omega = \omega_1) \\ = \; & \frac{1}{2} \times \frac{2}{5} \times \frac{2}{5} = 0.08 \end{aligned}$$

Similarly, for class $\omega = \omega_2$ we compute

$$\begin{aligned} & P(\omega_2) \prod_{i=1}^{2} P(A_i = a_i \mid \omega = \omega_2) \\ = \; & \frac{1}{2} \times \frac{1}{5} \times \frac{1}{5} = 0.02 \end{aligned}$$

The first probability for the class $\omega = \omega_1$ is higher. Therefore the class of the test sample x is predicted to be $\omega = \omega_1$. $\square$

Minimization of the classification error is not always the most satisfactory criterion to assign a data point to a particular class. Further, the Bayesian classification scheme assumes that all errors are of same significance. It is possible that some errors might have more weight than others. Therefore we can assign a penalty term to each error.