Exercise 1 - Conditional Probabilities

Data

Assume the data as shown below. The dataset is divided into 2 sectors.

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	-
3	0	1	1	-
4	0	1	1	-
5	0	0	1	+
6	1	0	1	+
7	1	0	1	-
8	1	0	1	-
9	1	1	1	+
10	1	0	1	+

Naive Bayes classifier

The fundamental assumption of Naive Bayes classifier is that each feature has *independent* and *equal* contribution to the outcome. For example, variable A has no effect to variable B or C and vice versa and each feature has the same *weight* at predicting. In order to calculate the following conditional probabilities. P(A/+), P(B/+), P(C/+), P(A/-), P(B/-), P(C/-), we need to isolate each feature and examine the probability of a record classified either in class + or -. We see from the data that there is an equal change something to be classified in each of 2 classes.

$$P(+) = \frac{1}{2}, P(-) = \frac{1}{2}$$

The probability of a record with A = 1 to be classified in class + or -, is the sum of all instances of A = 1 divided by total number of occurrences of class + or -. Thus we define the conditional probability of A = 1 given class + or - as follows.

$$P(A_{=1}|+) = \frac{3}{5}, P(A_{=1}|-) = \frac{2}{5}$$

Similarily, for A=0 given + or - is

$$P(A_{=0}|+) = \frac{2}{5}, P(A_{=0}|-) = \frac{3}{5}$$

Following the same approach we can calculate the probabilites for feature B and C.

$$P(B_{=1}|+) = \frac{1}{5}, P(B_{=1}|-) = \frac{2}{5}$$

$$P(B_{=0}|+) = \frac{4}{5}, P(B_{=0}|-) = \frac{3}{5}$$

$$P(C_{=1}|+) = \frac{4}{5}, P(C_{=1}|-) = \frac{5}{5}$$

$$P(C_{=0}|+) = \frac{1}{5}, P(C_{=0}|-) = \frac{0}{5}$$

Bayes theorem, measures the probability of an event Y occurring given some other event X is true/occurred. Mathematically is expressed as follow:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

In our case, Y is the class variable and X the feature vector. Given the independence among features assumption we made previously, we can a express the probability of occuring class Y given X vector as

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^{n} P(X_i|Y)}{P(X_1)P(X_2)....P(X_n)}$$

An estimate for class(+/-) for feature vector sample of X{A=0, B=1, C=0} is:

$$P(+|A_{=0}, B_{=1}, C_{=0}) = \frac{P(A_{=0}|+)P(B_{=}1|+)P(C_{=0}|+)P(+)}{P(A_{=0})P(B_{=1})P(C_{=0})}$$

$$P(+|A_{=0}, B_{=1}, C_{=0}) = \frac{\frac{2}{5}\frac{1}{5}\frac{1}{5}\frac{1}{2}}{\frac{1}{2}\frac{3}{10}\frac{1}{10}} = \frac{0.008}{0.015} \approx 0.53$$

and

$$P(-|A_{=0}, B_{=1}, C_{=0}) = \frac{P(A_{=0}|-)P(B_{=1}|-)P(C_{=0}|-)P(-)}{P(A_{=0})P(B_{=1})P(C_{=0})}$$

$$P(-|A_{=0}, B_{=1}, C_{=0}) = \frac{\frac{3}{5}\frac{2}{5}\frac{0}{5}\frac{1}{2}}{\frac{1}{2}\frac{3}{10}\frac{1}{10}} = 0$$

Since the demomintation is the same for both cases, we case simply get the highest enumerator.

$$Y = argmax_Y P(Y)\prod_{i=1}^{n} P(X_i|Y)$$

Thus, the sample X{A=0, B=1, C=0} is classified to class “+”.