Naive Bayes Classification

LJ Brown

May 3, 2018

Homework 6, Question 3

You need to categorize vehicles into the following categories: truck, suv, and sedan using 5 features. Perform a naïve Bayes classification with the following probabilities. Show your work.

Table 1

c	TRUCK	SUV	SEDAN
$P(c)$	0.35	0.4	0.25
$P(f_1\|c)$	0.2	0.01	0.2
$P(f_2\|c)$	0.01	0.1	0.05
$P(f_3\|c)$	0.1	0.001	0.005
$P(f_4\|c)$	0.001	0.2	0.005
$P(f_5\|c)$	0.005	0.008	0.01

Bayes' Theorem

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

Classifying c given a set of features

F is a set of features
f_i is a member of F

$C = \{\text{TRUCK, SUV, SEDAN}\}$
c_i is a member of C

For a given feature set, F, I classify c as,

$$c = \max_{c_i \in C} \big(P(c_i|F)\big)$$

Rewriting Bayes Theorem for a given class, c_i, and feature set, F,

$$P(c_i|F) = \frac{P(c_i)P(F|c_i)}{P(F)}$$

$$P(c_i|F) = \frac{P(c_i)P(F|c_i)}{P(F)} = P(c_i) \prod_{f_i \in F} \frac{P(f_i|c_i)}{P(f_i)}$$

When classifying c I think you can drop the denominator and just compare relative scores. I think this is fine because the events in C are mutually exclusive.

$$P(c_i) \prod_{f_i \in F} \frac{P(f_i|c_i)}{P(f_i)} \propto P(c_i) \prod_{f_i \in F} P(f_i|c_i)$$

$$P(c_i|F) \propto P(c_i) \prod_{f_i \in F} P(f_i|c_i)$$

$$\max_{c_i \in C} \big(P(c_i|F)\big) = \max_{c_i \in C} \Big(P(c_i) \prod_{f_i \in F} P(f_i|c_i)\Big)$$

So the final equation the program uses to classify is,

$$c = \max_{c_i \in C} \Big(P(c_i) \prod_{f_i \in F} P(f_i|c_i)\Big)$$

Where C and F are given as parameters.

Question 3 Answers

Question 3.a

a. What category would you assign to the vehicle (f1, f2, f3)?

$F = \{\text{f1, f2, f3}\}$

$$c = \max_{c_i \in C} \Big(P(c_i)P(f_1|c)P(f_2|c_i)P(f_3|c_i)\Big)$$

example calculation (happens to be the maximum) for a single value of c_i,

$$P(TRUCK|F) \approx 0.35 * 0.2 * 0.01 * 0.1 = 0.00007$$

$$c = \max_{c_i \in C} \Big(P(c_i)P(f_1|c)P(f_2|c_i)P(f_3|c_i)\Big) = \text{ TRUCK}$$

Answer for question 3.a is TRUCK.

Question 3.b

b. What category would you assign to the vehicle (f1, f2, f4, f5)?

$F = \{\text{f1, f2, f4, f5}\}$

$$c = \max_{c_i \in C} \Big(P(c_i)P(f_1|c)P(f_2|c_i)P(f_4|c_i)P(f_5|c_i) \Big) = \text{ SUV}$$

Answer for question 3.b is SUV.

Scratch

Below I'm trying to solve for $P(c_i|F)$, not sure if it works everytime.

F is a set of features
f_i is a member of F
c_i is a member of $C = \{\text{TRUCK, SUV, SEDAN}\}$

Ok so I'm assuming you can solve for $P(f_i)$ like this,

$$P(f_i) = \sum_{c_i \in C} P(f_i|c_i)$$

$P(f_1)$	0.41
$P(f_2)$	0.16
$P(f_3)$	0.106
$P(f_4)$	0.206
$P(f_5)$	0.023

Question 3.a

a. What category would you assign to the vehicle (f1, f2, f3)?

$$P(c|f_1 \cap f_2 \cap f_3) = \frac{P(c)P(f_1|c)P(f_2|c)P(f_3|c)}{P(f_1)P(f_2)P(f_3)}$$

$$P(\text{TRUCK}|f_1 \cap f_2 \cap f_3) = \frac{0.35 * 0.2 * 0.01 * 0.1}{0.41 * 0.16 * 0.106} \approx 0.0101$$

$$P(\text{SUV}|f_1 \cap f_2 \cap f_3) \approx 0.0005$$

$$P(\text{SEDAN}|f_1 \cap f_2 \cap f_3) \approx 0.0018$$

This gives the same answer for question 3.b: TRUCK.

Question 3.b

b. What category would you assign to the vehicle (f1, f2, f4, f5)?

$$P(\text{TRUCK}|f_1 \cap f_2 \cap f_4 \cap f_5) \approx 0.0001$$

$$P(\text{SUV}|f_1 \cap f_2 \cap f_4 \cap f_5) \approx 0.0021$$

$$P(\text{SEDAN}|f_1 \cap f_2 \cap f_4 \cap f_5) \approx 0.0004$$

This gives the same answer for question 3.b: SUV.