DJILLALI LIABES UNIVERSITY OF SIDI BEL ABBES
FACULTY OF EXACT SCIENCES
DEPARTMENT OF COMPUTER SCIENCES

Module : Apprentissage Automatique
1ST YEAR OF MASTER'S DEGEREE IN
NETWORKS,SYSTEMS & INFORMATION SECURITY(RSSI)
2021/2022

Solution To TD-03

Author:
HADJAZI Mohammed Hisham
Group: 01 / RSSI

Supervisor:
Pr.ELBERRICHI Zakaria

November 2, 2021

Contents

Chapter 1

Fiche TD-03 Solutions

1.1 Avez-vous refait tous les calculs du cours ? si oui continuez.

yes I did, but i keep getting different results every time I redo the exercise ..... Therefore I would appreciate it if we can get the solutions to compare it with our findings.

1.2 Pourquoi l'algorithme Naïve Bayes porte ce nom ?

The name consists of two parts, the first part **Naïve** is an English word that means natural and unaffected or innocent. which explains how this algorithm deals with the input data as it treats them equally without prior judgment (However in reality most data are related for example if the weather is cloudy the humidity will increase which means there is a relation between weather and humidity).

The second part of the name **Bayes** refers to the **Bayes Theorem** or **Bayes Theorem** the founder of the Bayes Theorem which is a probability theory which states

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

it came from the following probability formula :

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

where he substituted The upper side by the value of multiplying :

$$P(A \bigcap B) = P(B|A) * P(A)$$

1.3 Pourquoi l'estimateur de Laplace a été utilisé ?

Laplace smoothing in Naïve Bayes algorithm is used to tackle the problem of having a probability of **zero** in our calculations, the solution is very simple as we just add value of **one** to the nominator and dominator for example :

$$P(\omega|c) = \frac{Count(\omega,c)+\alpha}{Count(c)+\alpha} \text{ ,where } \alpha = 1$$

To conclude Laplace smoothing is a smoothing technique that helps tackle the problem of zero probability in the Naïve Bayes machine learning algorithm. Using higher alpha values will push the likelihood towards a value of 0.5, i.e., the probability of a word equal to 0.5 for both the positive and negative reviews. Since we are not getting much information from that, it is not preferable. Therefore, it is preferred to use alpha=1.[1]

1.4 Pourquoi les exemples traités au cours avaient des réponses très proches ?

Not sure I understood the question as we did only one instance in the cours.

1.5 Exercice 1 : Pour le tableau weather « symbolique » sans attributs numériques, trouver la classe des instances :

First we will make frequency tables with and without Laplace.

1.5.1 frequency tables without Laplace

Outlook	yes	no
Sunny	2/9	3/5
Overcast	4/9	0/5
Rainy	3/9	2/5
—	9/14	5/14

Temperature	yes	no
Hot	2/9	2/5
Mild	4/9	2/5
Cool	3/9	1/5
—	9/14	5/14

Humidity	yes	no
High	3/9	4/5
Normal	6/9	1/5
—	9/14	5/14

Windy	yes	no
True	3/9	3/5
False	6/9	2/5
—	9/14	5/14

1.5.2 frequency tables with Laplace

Outlook	yes	no
Sunny	3/12	4/8
Overcast	5/12	1/8
Rainy	4/12	3/8
—	9/14	5/14

Temperature	yes	no
Hot	3/12	3/8
Mild	5/12	3/8
Cool	4/12	2/8
—	9/14	5/14

Humidity	yes	no
High	4/11	5/7
Normal	7/11	2/7
—	9/14	5/14

Windy	yes	no
True	4/11	4/7
False	7/11	3/7
—	9/14	5/14

1.5.3 Overcast, hot, high, true

now solving This question without lapsing will yield the following :

$$P(Y/E) = 4/9 * 2/9 * 3/9 * 3/9 * 9/14 = 0.00705$$

$$P(N/E) = 0 becauese P(overcast/N) = 0$$

Since the probability of not playing with this instance **No = 0%** therefore the probability of playing will automaticity become **Yes = 100%**.

1.5.4 Sunny, ? , high, true

$$P(Y/E) = 2/9 * 3/9 * 3/9 * 9/14 = 0.015873 = 13.4\%$$

$$P(N/E) = 3/5 * 4/5 * 3/5 * 5/14 = 0.102857 = 86.6\%$$

As we can see **Play = No 86.6%**.

1.5.5 (rainy mild normal false) without Laplace

$$P(Y/E) = 3/9 * 4/9 * 6/9 * 6/9 * 9/14 = 0.042328 = 90.3\%$$

$$P(N/E) = 2/5 * 2/5 * 1/5 * 2/5 * 5/14 = 0.004571 = 9.7\%$$

Therefore **Play = yes 90.3%**.

1.5.6 (rainy mild normal false) with Laplace

$$P(Y/E) = 4/12 * 5/12 * 7/12 * 7/12 * 9/14 = 0.03616 = 85\%$$

$$P(N/E) = 3/8/3/8 * 2/7 * 3/7 * 5/14 = 0.00615 = 15\%$$

Therefore **Play = yes 85%**.

To my note I don't think that this is the best way to do smoothing as to my attention although the final result is the same the difference in percentage is huge. 5% is not small at all.

1.6 Exercice 2 : Cette fois pour le tableau weather «numérique » celui avec des attributs numériques, trouver la classe des instances :

Outlook	yes	no
Sunny	2/9	3/5
Overcast	4/9	0/5
Rainy	3/9	2/5
—	9/14	5/14

Temperature	yes	no
M	73	74.6
E	6.2	7.9

Humidity	yes	no
M	79.11	86.2
E	20.2	9.7

Windy	yes	no
True	3/9	3/5
False	6/9	2/5
—	9/14	5/14

1.6.1 Instance = rainy / 80 / 89 / true

$$P(80/Y) = \frac{1}{\sqrt{2\pi * 6.2}} * e^{-\frac{1}{2} * \frac{80-73^2}{6.2}} = 0.034$$

$$P(80/N) = \frac{1}{\sqrt{2\pi * 7.9}} * e^{-\frac{1}{2} * \frac{80-74.6^2}{7.9}} = 0.03998$$

$$P(89/Y) = \frac{1}{\sqrt{2\pi * 10.2}} * e^{-\frac{1}{2} * \frac{89-79.11^2}{10.2}} = 0.0244$$

$$P(89/N) = \frac{1}{\sqrt{2\pi * 9.7}} * e^{-\frac{1}{2} * \frac{89-86.2^2}{9.7}} = 0.0394$$

$$P(Y/E) = 3/9 * 0.034 * 0.0244 * 3/9 * 9/14 = 0.0000593 = 31\%$$

$$P(N/E) = 2/5 * 0.03998 * 0.0394 * 3/5 * 5/14 = 0.0001350 = 69\%$$

Therefore **Play = NO 69%**.

1.6.2 Instance = overcast / ? / 85 / false

$$P(Y/E) = 100\%$$

$$P(N/E) = 0\%$$

Because $P(Overcast/N) = 0$ then **Play = YES 100%**.

1.6.3 Instance = rainy / 86 / ? / false

$$P(Y/E) = 3/9 * 0.00714 * 6/9 * 9/14 = 0.00102 = 50\%$$

$$P(N/E) = 2/5 * 0.0178 * 2/5 * 5/14 = 0.00102 = 50\%$$

So **Play = yes 50%** and **Play = No 50%**.

1.7 Exercice 3 : En utilisant l'algorithme Naïve bayes, trouver la classe

1.7.1 frequency tables

Endroit	yes	no
B	2/9	3/5
R	4/9	0/5
V	3/9	2/5
—	9/14	5/14

Type	yes	no
U	2/9	2/4
J	4/9	1/4
R	3/9	1/4
—	9/14	5/14

Revenu	yes	no
M	50	75
E	27.26	24.24

Connu	yes	no
O	3/9	3/5
N	6/9	2/5
—	9/14	5/14

1.7.2 l'instance : V / ? / 68 / N

$$P(Y/E) = 3/9 * 0.0118 * 6/9 * 9/14 = 0.001686 = 65\%$$

$$P(N/E) = 2/5 * 0.0158 * 2/5 * 5/14 = 0.0009029 = 35\%$$

Therefore **Class = OUI 65%**.

Bibliography

[1] Vaibhav Jayaswal. *Laplace smoothing in Naïve Bayes algorithm*. Nov. 2020. URL: https://towardsdatascience.com/laplace-smoothing-in-na%C3%AFve-bayes-algorithm-9c237a8bdece.