Chair of Decision Sciences and Systems
TUM School of Computation, Information and Technology
Technical University of Munich

Business Analytics & Machine Learning
Tutorial sheet 4: Naïve Bayes – Solution

Prof. Dr. Martin Bichler
Julius Durmann, Markus Ewert, Yutong Chao, Dr. Mete Ahunbay

Exercise T4.1 *Cookie factory*

You are the manager of a company which produces cookies and you want to introduce a new product. Your R&D department has proposed and developed the following two alternatives:

1. Unicorn cookies (UC)
2. Vanilla-chip cookies (VC).

As part of your market research, you are interested in predicting whether certain customers are likely to buy one of the new products. For that, you have already collected data from a large number of test persons. In particular, you asked them to fill out a query with the following questions:

1. *What do you think is the most fascinating: Rainbows, Black holes or Cats?* (variable $preferences$)
2. *How much money do you spend on cookies per month?* (variable $money$)
3. *Which of our cookies would you buy?* (variable $product$)
   Note: The variable $product$ can also take on the value "No product" (NP).

You can find the data in *cookie-factory.csv*. We recommend using the provided notebook template to solve sub-tasks b) - e).

a) For each of the questions 1-3, decide
   (i) whether the answers are continuous or discrete outcomes,
   (ii) which range the outcomes could have

To infer which products new customers are likely to buy, you set up a probabilistic model. You assume that the answers to questions 1 - 2 are conditionally independent (Naive Bayes) given $product$ and model the dependencies as follows:

$$f(preferences, money, product) = \mathbb{P}(preferences \mid product) \cdot f_{money}(money \mid product) \cdot \mathbb{P}(product)$$

b) Estimate the parameters of your categorical prior $\mathbb{P}(product)$ by using maximum likelihood:

$$\mathbb{P}(product = UC) = p_{UC} \qquad \mathbb{P}(product = VC) = p_{VC} \qquad \mathbb{P}(product = NP) = p_{NP}$$

   Hint: The maximum likelihood estimate of the parameters for categorically distributed variables is simply the fraction of samples from a category.

Based on your observations in a), you decide to model the likelihoods as follows:

1. $preferences$ follows a Categorical distribution where the parameters depend on the product the customers would buy:

$$\begin{aligned}
&\mathbb{P}(preferences = \text{"Rainbows"} &&\mid prod. = \text{UC}) = \pi_R^{UC}\\
&\mathbb{P}(preferences = \text{"Black holes"} &&\mid prod. = \text{UC}) = \pi_B^{UC}\\
&\mathbb{P}(preferences = \text{"Cats"} &&\mid prod. = \text{UC}) = \pi_C^{UC}\\
\\
&\mathbb{P}(preferences = \text{"Rainbows"} &&\mid prod. = \text{VC}) = \pi_R^{VC}\\
&\mathbb{P}(preferences = \text{"Black holes"} &&\mid prod. = \text{VC}) = \pi_B^{VC}\\
&\mathbb{P}(preferences = \text{"Cats"} &&\mid prod. = \text{VC}) = \pi_C^{VC}\\
\\
&\mathbb{P}(preferences = \text{"Rainbows"} &&\mid prod. = \text{NP}) = \pi_R^{NP}\\
&\mathbb{P}(preferences = \text{"Black holes"} &&\mid prod. = \text{NP}) = \pi_B^{NP}\\
&\mathbb{P}(preferences = \text{"Cats"} &&\mid prod. = \text{NP}) = \pi_C^{NP}
\end{aligned}$$

2. $money$ follows an exponential distribution where the parameter $\lambda_{product}$ depends on the product the customers would buy ($\eta_{product} = \eta_{UC}$, $\eta_{product} = \eta_{VC}$ or $\eta_{product} = \eta_{NP}$):

$$f_{money}(m|product) = \begin{cases} \eta_{product} \cdot e^{-\eta_{product} \cdot m} & m \geq 0 \\ 0 & \text{else} \end{cases}$$

Intuitively, your model describes the profile ($preferences$, $money$) of a customer if you already know which product they would buy ($product$).

c) Using the data, derive maximum likelihood estimates for all parameters.

Hint: The maximum likelihood estimate of the parameters for exponentially distributed variables is the inverse of their sample mean: $\bar{x}^{-1}$.

You now have access to a joint density over your data:

$$f(preferences, money, product) = \mathbb{P}(preferences \mid product) \cdot f_{money}(money \mid product) \cdot \mathbb{P}(product)$$

d) With the fitted model, predict the (posterior) probability

$$\mathbb{P}(product \mid preferences, money)$$

that the customers below buy a unicorn cookie, a vanilla-chip cookie or no cookie at all:

Customer	$preferences$	$money$
Anna	Cats	53.10 €
Ben	Rainbows	2.30 €
Caroline	Black holes	10.25 €

e) From a fourth customer, you only know that they like rainbows. Predict the probability that they buy unicorn cookies.

f) *[Bonus]* You may have noticed that the data also contains information about age. What would you need to do to include this information as well?

Solution

The solution below is meant as a reference. Please also have a look at the provided solution notebook which should give more insight into the computation.

a) The table below shows our suggestion. Some of the answers could surely be discussed. For instance, the variable $money$ could be considered discrete since cookie prices can only differ up to a unit of 1 cent.

Variable	Discrete?	Range
$money$	continuous	$[0, \infty)$
$preferences$	discrete	$\{"rainbows", "black\ holes", "cats"\}$
$product$	discrete	$\{"UC", "VC", "NP"\}$

b) The optimal parameters are as follows:

p_{UC}	0.613
p_{VC}	0.199
p_{NP}	0.188

c) The optimal parameters are as follows:

1.

π_R^{UC}	0.5987	π_B^{UC}	0.3083	π_C^{UC}	0.0930
π_R^{VC}	0.4255	π_B^{VC}	0.1755	π_C^{VC}	0.3990
π_R^{NP}	0.3920	π_B^{NP}	0.3819	π_C^{NP}	0.2261

2.

η_{UC}	0.0913
η_{VC}	0.0664
η_{NP}	0.1122

d) With the joint distribution we can derive the posterior:

$$\begin{aligned}
&\mathbb{P}(product \mid money, preferences) \\
&\quad = \frac{f(preferences, money, product)}{f(preferences, money)} \\
&\quad = \frac{1}{Z} \cdot f(preferences, money, product) \\
&\quad = \frac{1}{Z} \cdot \mathbb{P}(preferences \mid product) \cdot f_{money}(money \mid product) \cdot \mathbb{P}(product)
\end{aligned}$$

By plugging in the values from b) and c), we can compute the individual terms. Once this is achieved for all products, the normalizing constant Z is such that the posterior probabilities over all products (UC, VC, NP) sum up to 1.

For the customer Anna, we compute the *unnormalized* posterior values:

$$\begin{aligned}
&\tilde{\mathbb{P}}(product = UC \mid money = 53.10€,\ preferences = "Cats") \\
&\qquad = \mathbb{P}(preferences = "Cats" \mid UC) \cdot f_{money}(53.10 \mid UC) \cdot \mathbb{P}(product = UC) \\
&\tilde{\mathbb{P}}(product = VC \mid money = 53.10€,\ preferences = "Cats") \\
&\qquad = \mathbb{P}(preferences = "Cats" \mid VC) \cdot f_{money}(53.10 \mid VC) \cdot \mathbb{P}(product = VC) \\
&\tilde{\mathbb{P}}(product = NP \mid money = 53.10€,\ preferences = "Cats") \\
&\qquad = \mathbb{P}(preferences = "Cats" \mid NP) \cdot f_{money}(53.10 \mid NP) \cdot \mathbb{P}(product = NP)
\end{aligned}$$

We find the normalizing constant Z as

$$\begin{aligned}
Z = \quad & \tilde{\mathbb{P}}(product = UC \mid money = 53.10€,\ preferences = "Cats") \\
+ \quad & \tilde{\mathbb{P}}(product = VC \mid money = 53.10€,\ preferences = "Cats") \\
+ \quad & \tilde{\mathbb{P}}(product = NP \mid money = 53.10€,\ preferences = "Cats")
\end{aligned}$$

Finally, we can use Z to normalize, and obtain:

$$\begin{aligned}
&\mathbb{P}(product = UC \mid money = 53.10€,\ preferences = "Cats") \\
&\qquad = \frac{1}{Z} \cdot \tilde{\mathbb{P}}(product = UC \mid money = 53.10€,\ preferences = "Cats") \\
&\mathbb{P}(product = VC \mid money = 53.10€,\ preferences = "Cats") \\
&\qquad = \frac{1}{Z} \cdot \tilde{\mathbb{P}}(product = VC \mid money = 53.10€,\ preferences = "Cats") \\
&\mathbb{P}(product = NP \mid money = 53.10€,\ preferences = "Cats") \\
&\qquad = \frac{1}{Z} \cdot \tilde{\mathbb{P}}(product = NP \mid money = 53.10€,\ preferences = "Cats")
\end{aligned}$$

For the other customer profiles, the formulas are similar.

The computation could be done by hand. Since this is tedious, we rely on Python and suggest you have a look at the provided solution notebook.

The resulting (rounded) probabilities are as follows:

	No product	Unicorn cookie	Vanilla cookie
Anna	0.0653	0.2036	0.7311
Ben	0.1756	0.7059	0.1185
Caroline	0.2552	0.6399	0.1049

e) If we have only limited information, we must consider the marginal joint distribution

$$\mathbb{P}(preferences, product) = \mathbb{P}(preferences \mid product) \cdot \mathbb{P}(product)$$

The posterior becomes

$$\mathbb{P}(product \mid preferences) = \frac{1}{Z} \cdot \mathbb{P}(preferences \mid product) \cdot \mathbb{P}(product)$$

Plugging in the maximum-likelihood values and normalizing gives

	No product	Unicorn cookie	Vanilla cookie
Unknown customer	0.1486	0.6990	0.1524

f) If we want to include information about age as well, we need to do the following:

1. Determine a suitable likelihood function form. For the age this could be a normal distribution or a Poisson distribution, depending on the assumptions that we make.
2. Derive estimates for the parameters of the likelihood distribution. As before, we need to find suitable parameterizations for our new distribution. Analogously to before, we could do this by using maximum likelihood estimates.
3. Compute the posterior probability using Bayes rule. This is similar to before, but we add a new term to the *joint* distribution before computing the posterior:

$$f(\mathbf{age}, preferences, money, product) = \\ \mathbf{f}(\mathbf{age} \mid \mathbf{product}) \cdot \mathbb{P}(preferences \mid product) \cdot f_{money}(money \mid product) \cdot \mathbb{P}(product)$$

The posterior becomes

$$f(product \mid \mathbf{age}, preferences, money) = \frac{1}{Z} \cdot f(\mathbf{age}, preferences, money, product)$$

Exercise T4.2 *Sushi or Pizza*

Anna and Ben want to have dinner together, but they cannot decide if they prefer pizza or sushi. For this reason, they design the following mechanism to make a decision.

1. They both roll a dice (fair, six-sided), obtaining events A (Anna's dice) and B (Ben's dice).
2. They determine the sum of the outcomes and note down if it is even or odd. (Event $C \in \{"e", "o"\}$)
3. They throw a biased coin which shows head with 80% probability and tails with 20%. If $C = "e"$, they opt for pizza ($"P"$) if head is shown and sushi ($"S"$) if tails is shown. If $C = "o"$, they opt for sushi if head is shown and pizza if tails is shown. (Event $D \in \{"P", "S"\}$)

a) Visualize the process as a Bayesian network.

b) Next to the nodes of the Bayesian network, note the likelihood tables of the conditional probabilities.

c) Write down the joint probability distribution in terms of A, B, C, and D.

d) Compute the probability that they decide on pizza, given that Anna's dice shows 2 and Ben's dice shows 3.

e) Are A and B independent? Are A and B *conditionally* independent given C?

Solution

a) This is what the Bayesian network should look like:

$\mathbb{P}(A)$

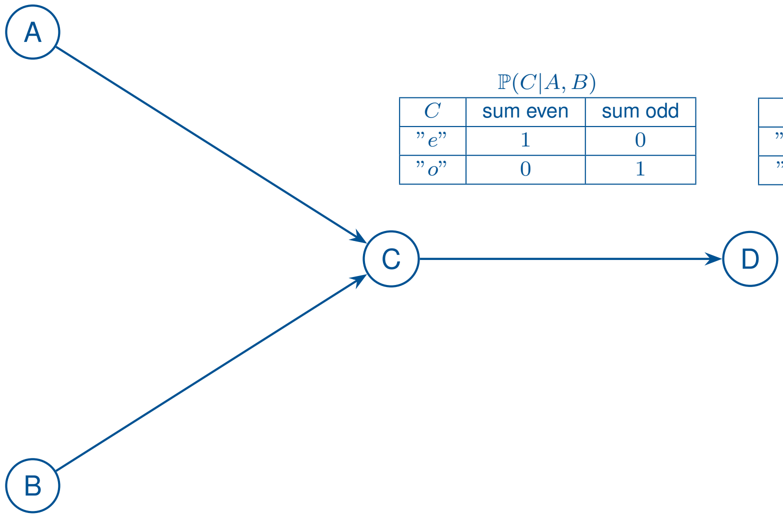
A	
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$\mathbb{P}(C|A,B)$

C	sum even	sum odd
"e"	1	0
"o"	0	1

$\mathbb{P}(D|C)$

D	$C=$"e"	$C=$"o"
"P"	0.8	0.2
"S"	0.2	0.8



$\mathbb{P}(B)$

B	
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

b) (See above)

c)

$$\mathbb{P}(A,B,C,D) = \mathbb{P}(A)\cdot\mathbb{P}(B)\cdot\mathbb{P}(C|A,B)\cdot\mathbb{P}(D|C)$$

d)

$$\begin{aligned}
\mathbb{P}(D={}"P"|A=2,B=3) &= \sum_{c\in\{"e","o"\}} \mathbb{P}(C=c,D={}"P"|A=2,B=3)\\
&= \sum_{c\in\{"e","o"\}} \mathbb{P}(C|A=2,B=3)\cdot\mathbb{P}(D={}"P"|C=c,A=2,B=3)\\
&= \sum_{c\in\{"e","o"\}} \mathbb{P}(C|A=2,B=3)\cdot\mathbb{P}(D={}"P"|C=c)\\
&= \underbrace{0\cdot 0.8}_{c="e"}+\underbrace{1\cdot 0.2}_{c="o"}\\
&= 0.2
\end{aligned}$$

e) The two variables are independent by construction, and we find $\mathbb{P}(A=a,B=b)=\mathbb{P}(A=a)\cdot\mathbb{P}(B=b)$.

To answer if they are conditionally independent, we must check if, for all a, b, and c

$$\mathbb{P}(A=a,B=b|C=c) \stackrel{?}{=} \mathbb{P}(A=a|C=c)\cdot\mathbb{P}(B=b|C=c) \tag{1}$$

We first evaluate the left side of equation (1), starting with the joint probability of A, B, and C.

$$\begin{aligned}
\mathbb{P}(A=a,B=b,C={}"o") &= \mathbb{P}(A=a)\cdot\mathbb{P}(B=b)\cdot\mathbb{P}(C={}"o"|A=a,B=b)\\
&= \frac{1}{6}\cdot\frac{1}{6}\cdot\begin{cases}1 & \text{if } a+b \text{ is odd}\\ 0 & \text{else}\end{cases}
\end{aligned}$$

Also, we will need $\mathbb{P}(C = "o")$. Since there are as many cases where $A+B$ are odd as there are cases where they are even, and since all cases are equally likely, $\mathbb{P}(C = "o") = \mathbb{P}(C = "e") = \frac{1}{2}$. With that, we can compute the conditional probability of A and B, given $C = "o"$.

$$\begin{aligned}
\mathbb{P}(A = a, B = b | C = "o") &= \frac{\mathbb{P}(A = a, B = b, C = "o")}{\mathbb{P}(C = "o")} \\
&= \frac{1}{18} \cdot \begin{cases} 1 & \text{if } a+b \text{ is odd} \\ 0 & \text{else} \end{cases}
\end{aligned}$$

To evaluate the right side of equation (1), we marginalize over B.

$$\begin{aligned}
\mathbb{P}(A = a | C = "o") &= \sum_{b=1}^{6} \mathbb{P}(A = a, B = b | C = "o") \\
&= \sum_{b=1}^{6} \frac{1}{18} \cdot \begin{cases} 1 & \text{if } a+b \text{ is odd} \\ 0 & \text{else} \end{cases} \\
&= 3 \cdot \frac{1}{18} = \frac{1}{6}
\end{aligned}$$

In the last line, we used that, no matter what a is, there are always three cases for B such that $a+b$ is odd. Similarly, by marginalizing over A, we find for B: $\mathbb{P}(B = b | C = "o") = \frac{1}{6}$. Fun fact: By comparing $\mathbb{P}(A)$ with $\mathbb{P}(A | C = "o")$, we find that A is independent of C. (Pairwise independence does not indicate independence of multiple random variables.)

We insert the computed values into equation (1):

$$\mathbb{P}(A = a, B = b | C = "o") = \frac{1}{18} \cdot \begin{cases} 1 & \text{if } a+b \text{ is odd} \\ 0 & \text{else} \end{cases} \neq \frac{1}{36} = \mathbb{P}(A = a | C = "o") \cdot \mathbb{P}(B = b | C = "o")$$

This means that the two variables A and B are not conditionally independent given C.