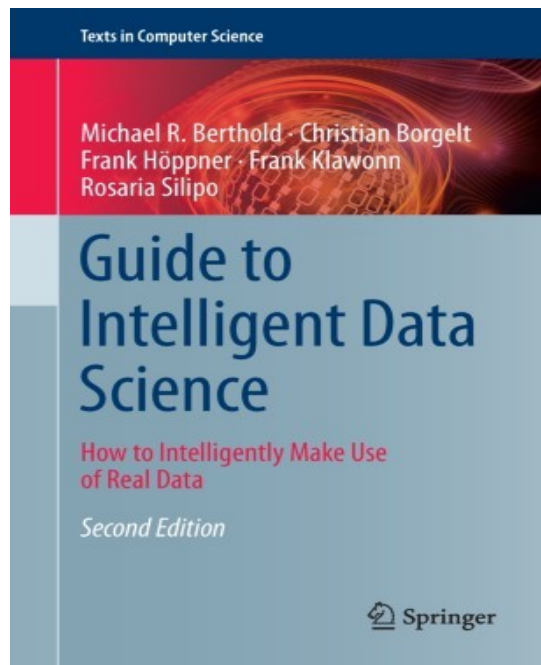




Big Data ACUA6002 Tutorial 7

Bayes Classifiers

Exercise 1

Bayes Theorem

- The conditional probability , hypothesis is true given event

$$P(h|e) = \frac{P(e|h) \cdot P(h)}{P(e)}$$

- : Probability of hypothesis
- : Probability of event
- : Conditional probability of event given hypothesis

Answer the following questions and give an explanation of your answers.

- 1. 2.6% of the total population is color-blind. On average, 5 out of 100 men and 25 out of 10000 women are color-blind. We assume that the number of men and women in the total population is equivalent. Now we choose a color-blind person from the population, what is the probability that this person is a man?*
- 2. In a company, machine M1 produces 20% of the total production, machine M2 produces 30% and machine M3 produces 50%. On average, 2% of the production of M1, 4% of M2 and 7% of M3 is defective. We take out one defective part of the production, what is the probability that it has been produced by M1/M2/M3?*
- 3. Your mailbox contains 100 unread emails. 80% of these mails are spam, 20% are non-spam. 25% of spam emails and 50% of non-spam emails contain the word 'urgent'. You open a random email and it contains the word 'urgent'. What is the probability that this is a spam email? What is the probability that this is not a spam email?*

1. *2.6% of the total population is color-blind. On average, 5 out of 100 men and 25 out of 10000 women are color-blind. We assume that the number of men and women in the total population is equivalent. Now we choose a color-blind person from the population, what is the probability that this person is a man?*
- Let F be the event that the person is a woman and M the event that the person is a man. B is the event that the person is color-blind. The following probabilities are given:
 - Inserting those values into the Bayes Theorem we get:

2. *In a company, machine M1 produces 20% of the total production, machine M2 produces 30% and machine M3 produces 50%. On average, 2% of the production of M1, 4% of M2 and 7% of M3 is defective. We take out one defective part of the production, what is the probability that it has been produced by M1/M2/M3?*
- Let A be the event that the part has been produced by machine $M1$, B the event that the part is defective. The following probabilities are given:
 - To use Bayes theorem we need to calculate the total probability for :

2. *In a company, machine M1 produces 20% of the total production, machine M2 produces 30% and machine M3 produces 50%. On average, 2% of the production of M1, 4% of M2 and 7% of M3 is defective. We take out one defective part of the production, what is the probability that it has been produced by M1/M2/M3?*

– On average 5.1% of all parts are defective. With Bayes Theorem:

—

—

—

3. *Your mailbox contains 100 unread emails. 80% of these mails are spam, 20% are non-spam. 25% of spam emails and 50% of non-spam emails contain the word 'urgent'. You open a random email and it contains the word 'urgent'. What is the probability that this is a spam email? What is the probability that this is not spam?*
- Let A be the event that the email is spam and B the event that the email is not spam. C is the event that the email contains the word “urgent”. The following probabilities are given:
 - To use the Bayes Theorem we need to calculate the total probability for:

3. *Your mailbox contains 100 unread emails. 80% of these mails are spam, 20% are non-spam. 25% of spam emails and 50% of non-spam emails contain the word 'urgent'. You open a random email and it contains the word 'urgent'. What is the probability that this is a spam email? What is the probability that this is not spam?*

– With Bayes Theorem:

Exercise 2

Naïve Bayes

Naïve Bayes

- *The following database for fitting contact lenses is given. Based on the attributes age, spectacle prescription, astigmatism and tear production rate the task is to predict whether a patient should be fitted with hard contact lenses, soft contact lenses or not be fitted with contact lenses at all.*

age	spec-prescrib	astigmatism	tear-prod-rate	contact lenses
young	myope	no	reduced	none
young	myope	no	normal	soft
young	myope	yes	reduced	none
young	myope	yes	normal	hard
young	hypermetrope	no	reduced	none
young	hypermetrope	no	normal	soft
young	hypermetrope	yes	normal	hard
pre-presbyopic	myope	no	reduced	none
pre-presbyopic	myope	yes	normal	hard
pre-presbyopic	hypermetrope	no	reduced	none
pre-presbyopic	hypermetrope	no	normal	soft
pre-presbyopic	hypermetrope	yes	reduced	none
pre-presbyopic	hypermetrope	yes	normal	none
presbyopic	myope	no	reduced	none
presbyopic	myope	no	normal	none
presbyopic	myope	yes	reduced	none
presbyopic	myope	yes	normal	hard
presbyopic	hypermetrope	no	reduced	none
presbyopic	hypermetrope	no	normal	soft
presbyopic	hypermetrope	yes	reduced	none
presbyopic	hypermetrope	yes	normal	none

- *Compute all probabilities that are necessary to predict the following instances with a Naïve Bayes classifier and classify them*

age	spec-prescrib	astigmatism	tear-prod-rate
young	hypermetrope	yes	reduced
pre-presbyopic	myope	no	normal
pre-presbyopic	myope	yes	reduced

Naïve Bayes

- The conditional probability tables for all attributes:

Class	hard	none	soft
Count	4	13	4

Class/age	pre-presbyopic	presbyopic	young
hard	1	1	2
none	4	6	3
soft	1	1	2

Class/astig.	no	yes
hard	0	4
none	7	6
soft	4	0

Class/tear-prod-rate	normal	reduced
hard	4	0
none	3	10
soft	4	0

Class/spec-prescrip	hypermetrope	myope
hard	3	1
none	7	6
soft	3	1

Naïve Bayes

age	spec-prescrib	astigmatism	tear-prod-rate
young	hypermetrope	yes	reduced

- hard
- none
- soft

Naïve Bayes

age	spec-prescrib	astigmatism	tear-prod-rate
pre-presbyopic	myope	no	normal

- hard
- none
- soft

Naïve Bayes

age	spec-prescrib	astigmatism	tear-prod-rate
pre-presbyopic	myope	yes	reduced

- hard
- none
- soft

Berthold, M.R., Borgelt, C., Höppner, F., Klawonn, F. and Silipo, R. (2020). *Guide to Intelligent Data Science: How to Intelligently Make Use of Real Data*. 2nd ed. [online] www.springer.com. Springer International Publishing.

Available at:

<https://www.springer.com/gp/book/9783030455736> [Accessed 17 Jun. 2021].

Thank you

For any questions please contact: education@knime.com