

Techniques of Artificial Intelligence

Exercises – Naïve Bayes Classifier & WEKA

Dipankar Sengupta
Dipankar.Sengupta@vub.ac.be

Roxana Rădulescu
Roxana.Radulescu@vub.ac.be

April 18, 2016

23. Bayes theorem

For the course X, we experienced that on average one out of ten students passes. We also noticed over the last couple of years that from all the students who passed, 90% did attend the exercise sessions. From all the students who did not pass, 95% did not attend the exercise sessions; they preferred to go to the university pub. Are your chances for passing course X increased by attending the exercise sessions?

24. Bayes theorem

An HIV test gives a positive result with probability 98% when the patient is indeed affected by HIV, while it gives a negative result with 99% probability when the patient is not affected by HIV. If a patient is drawn at random from a population in which 0.1% of individuals are affected by HIV and he is found positive, what is the probability that he is indeed affected by HIV?

25. Naïve Bayes classifier

At the parking lot of company X, a lot of cars get stolen. See below for an overview of the last 10 cars which were parked. I now park my brand new RED DOMESTIC SUV, what is the maximum a posteriori hypothesis (MAP): will the car be stolen or not according to a naïve bayes classifier?

Example Number	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

26. Getting to know WEKA

Using WEKA you can load datasets, process them and apply classifiers on them. Open the “weather.nominal.arff” dataset, inspect it and run the ID3 algorithm on it. What is the performance of the classifier?

27. Explorer (Naïve Bayes and cross validation)

Use “blood_fat_corrupted.arff” dataset.

- How many attributes and instance are there? What are the possible categorical values?
- Identify the corrupted values. (instance and attributes)
- Rectify the corrupted values (ask the instructor for correct values).
- Apply a Naïve Bayes classifier with 5-fold cross validation, increment it by 5 and iterate it till 40-fold. Describe and motivate observed trend in the performance.

28. Naïve Bayes in WEKA

Open the “splice.arff” dataset.

- Investigate the performance of the Naïve Bayes learner.
- The dataset contains 3190 instances. Successively apply the “RemovePercentage” filter, such that the dataset size is reduced successively up to 50 instances. At each step, investigate the performance of the classifier. Describe and motivate the observed trend in the performance.

29. ID3 and WEKA

- Using the same dataset, show that ID3 does not manage to classify the data. Why is that? Investigate the data in the editor (preprocess tab).
- You can solve the problem by applying the correct filter on the data. (preprocess tab, choose an unsupervised attribute filter). Show that performance is increased in an important manner.