

COMP3308/3608 Artificial Intelligence

Weeks 6 Tutorial exercises Naïve Bayes. Evaluating Classifiers.

Exercise 1. Naïve Bayes (Homework)

Suppose you want to recognize good and bad items produced by your company. You are able to measure two properties of each item (P1 and P2) and express them with Boolean values. You randomly grab several items and test if they are good or bad, obtaining the following results:

P1	P2	result
Y	Y	good
Y	N	bad
N	N	good
N	Y	bad
Y	N	good
N	N	good

Use Naïve Bayes to predict the class, *good* or *bad*, of the following new item: P1=N, P2=Y. If there are ties, make a random choice.

Exercise 2. Naïve Bayes

Why is the Naïve Bayesian classification called “naïve”?

Exercise 3. Applying Naïve Bayes to data with both numerical and nominal attributes

Given is the training data in the table below (the *weather* data with some numerical attributes, *play* is the class). Predict the class of the following new example using the Naïve Bayes classification:

outlook=overcast, temperature=60, humidity=62, windy=false.

outlook	temperature	humidity	windy	play
sunny	85	85	false	no
overcast	80	90	true	no
overcast	83	86	false	yes
rainy	70	96	false	yes
rainy	68	80	false	yes
rainy	65	70	true	no
overcast	64	65	true	yes
sunny	72	95	false	no
sunny	69	70	false	yes
rainy	75	80	false	yes
sunny	75	70	true	yes
overcast	72	90	true	yes
overcast	81	75	false	yes
rainy	71	91	true	no

Exercise 4. Bayes Theorem (Advanced only)

Suppose that the fraction of undergraduate students who smoke is 15% and the fraction of graduate students who smoke is 23%. If 1/5 of the University students are graduate students and the rest are undergraduates, what's the probability that a student who smokes is a graduate student?

Hint: Use the Bayes Theorem; you will need to calculate the denominator using the *law of total probability*, see its Wikipedia description.

Exercise 5. Using Weka – Comparing Classifiers

1. Load the iris dataset
2. Choose “Percentage split” mode for evaluation: 66% training set, 33% testing set
3. Run the Naïve Bayes and review Weka's output
4. For comparison, also run k-nearest neighbor with k=1 and 3 (IB1 and IBk), OneR and ZeroR. Which is the most accurate classifier?
5. Change the test mode to “Cross validation”. Apply 10-fold cross validation instead of percentage split as evaluation mode and compare the classifiers.

Which classifier produced the most accurate classification?

Which evaluation strategy (percentage split or 10-fold cross validation) produced better results?

Which evaluation strategy, percentage split or cross validation, is more statistically reliable and why?

6. Apply leave-one-out cross validation. Tip: You need to specify the number of folds in the WEKA's cross validation box.
7. Check the confusion matrix printed by WEKA for one of the classifiers, e.g. Naïve Bayes, and verify the accuracy, recall, precision and F1 measure. Note: Weka shows recall, precision and F1 for each class separately.

Additional exercises to be done at your own time:**Exercise 6. Naïve Bayes with Laplace correction**

As in exercise 1, but now suppose that you are able to measure 3 properties of each item (P1, P2 and P3) and the data is as follows:

P1	P2	P3	result
Y	Y	Y	good
Y	N	N	bad
N	N	Y	good
N	Y	N	bad
Y	N	Y	good
N	N	N	good

Use Naïve Bayes to predict the class of the following new example P1=N, P2=Y P3=Y. If necessary use the Laplace correction.