

**Data Mining, CSCI 347, Fall 2019**  
**Homework 1, Naïve Bayes, due Sept. 23**

Consider the following database with the class value being “buys\_computer”.  
(In this image the open parenthesis represents a less-than sign, so the “(<=30” value for age represents “<=30”.)

Relation: Customers					
No.	buys_computer Nominal	age Nominal	income Nominal	student Nominal	credit_rating Nominal
1	no	(<=30	high	no	fair
2	no	(<=30	high	no	excellent
3	yes	31..40	high	no	fair
4	yes	)40	medium	no	fair
5	yes	)40	low	yes	fair
6	no	)40	low	yes	excellent
7	yes	31..40	low	yes	excellent
8	no	(<=30	medium	no	fair
9	yes	(<=30	low	yes	fair
10	yes	)40	medium	yes	fair
11	yes	(<=30	medium	yes	excellent
12	yes	31..40	medium	no	excellent
13	yes	31..40	high	yes	fair
14	no	)40	medium	no	excellent
15	yes	)40	high	no	excellent

- Write the formula to predict if a 25 year old student with a low income and a fair credit rating is likely to purchase a computer using Naïve Bayes Theorem. That is, write formulas for the following.

$$\Pr[\text{buys\_computer}=\text{'yes'} \mid \text{age is } \leq 30 \ \& \ \text{income}=\text{'low'} \ \& \ \text{student}=\text{'yes'} \ \& \ \text{credit\_rating}=\text{'fair'}]$$

$$\Pr[\text{buys\_computer}=\text{'no'} \mid \text{age is } \leq 30 \ \& \ \text{income}=\text{'low'} \ \& \ \text{student}=\text{'yes'} \ \& \ \text{credit\_rating}=\text{'fair'}]$$

2. Since this dataset is small, determine the values to use in the above formulas, counting instance by hand. Apply a Laplace estimator of 1, to avoid probabilities of 0.

3. Calculate the values, ignoring the denominators.

4. Normalize the results.

5. What would be predicted?

6. Use Weka to check your results.
  - a. Go to the classify tab and make “buys\_computer” the class attribute (can be set using the drop-down box that appears above the “Start” button)
  - b. Run the classifier weka.classifiers.bayes,NaiveBayes on the dataset (you can use cross-validation or use the training set for testing because the classifier model uses the full training set)
  - c. Compare Weka’s statistical values with yours.

6. Use Weka to classify the above instance.

buys_computer	age	income	student	credit_rating
?	<=30	low	yes	fair

In order to do this do the following:

- a. Create a new dataset which is like the original, only it just contains the above instance.

Your dataset can be:

```
% Test dataset consisting of a single instance

@relation Customers

@attribute buys_computer {yes, no}
@attribute age {<=30, 31..40, >40}
@attribute income {high, medium, low}
@attribute student {yes, no}
@attribute credit_rating {fair excellent}

@data
?,<=30, low, yes, fair
```

- b. On the Classify tab, in the Test options area, chose “Supplied test set” and select the new dataset with the single instance.
- c. Click the “more options...” button and check the “Output predictions” box.
- d. Run the Naïve Bayes classifier to see how Weka would classify this instance.
- e. Do the results match what you expected.