HARRISBURG UNIVERSITY
OF SCIENCE AND TECHNOLOGY

CISC 340

Introduction to Artificial Intelligence ¶

Lab Booklet #04

Lab Title: Supervised Learning

Instructors

Brian Grey M.S.

Chad Van Chu M.S.

Objectives

Supervised learning is the "entry level" approach to machine learning. In order to effectively do supervised learning, we need a data set which has already been labeled for us and an approach we need to take. In this lab, we will be examing two different approaches to supervised learning. We will be implementing a probablisitic approach using the Naive Bayes classifier and a decision tree approch using the ID3 learning algorithm.

After completing this lab, you will be able to:

- **Understand how to run the calculations for a Naive Bayes classifier**
- **Understand how to implement the ID3 learning algorithm**
- **Be able to compare and contrast the output from each approach**

Resources

While the lab will explain everything that you need to know, you may want to consult the following sources for additional knowledge:

- https://en.wikipedia.org/wiki/Naive_Bayes_classifier (https://en.wikipedia.org/wiki/Naive_Bayes_classifier)
- https://en.wikipedia.org/wiki/ID3_algorithm (https://en.wikipedia.org/wiki/ID3_algorithm)

Deliverables

For this lab, you will need to submit:

- Jupyter notebook files (.ipynb), named and coded as instructed.
  - CISC 340 Lab 04 FA19.ipynb
  - Excel Spreadsheet if you choose to use it for PART 1.

Instructions & Questions

We have provided a data set for this lab, `<your name> Lab04Data.xlsx` , that you will need to use for both parts of the lab. You can find this data file under the Lab 4 drop box on Moodle as a feedback file. Part 1 will involve running calculations in Excel. **Be sure NOT to alter the underlying data and be sure that your data is unaltered for Part 2.**

Part 1: Naive Bayes Classifier

Naive bayes classifiers work with Bayes' Theorem. Where we can calculate the posterior probability of an event occuring given that some other event has occured from the likelihood, prior probability, and normalizing constant (evidence) of those events.

Bayes' theorem

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

Before constructing any model, let's begin by opening the data file provided to you in the instructions to familiarize ourselves with the data we are going to be working with. The class label in our datasets will always be the left most column, in this case **play**.

1. What are the features that are present? How many records do you have present?

Type answer here!

2. Using the dataset, calculate the probablities of each of your present features given the class and also the probablilty of just the class. You may do this using a program or using just the excel spreadsheet provided. Whichever method you choose, be sure that the probabilities are clearly shown in order to get credit.

Type answer here!

3. Using your now "trained" model from Question 2, use those calculations to calculate what class the following new records would have. Show your calculations. (Be sure to not recalculate your probabilities, you already did that in part 2):

Outlook	Temperature	Humidity	Windy	Play
Sunny	Mild	Normal	False	---
Overcast	Hot	Normal	True	---
Rainy	Hot	High	True	---

Note: Feel free to edit this Markdown and replace the "---" with the actual class. But also remember you MUST include your calculations through provided code OR spreadsheet.

In [6]:
```
#Write code here if you wish to use programming instead of Excel.
'''Extra Credit will be awarded if you choose to program the
    calculations instead of using the excel spreadsheet.'''
```

Part 2: ID3 Decision Tree

4. Create an object called `ID3` that will contain your trained classifier. This object should have 3 publicly visible methods:

- **`train(rawData)` : Method which takes the raw data as a parameter and trains the decision tree.**
- **`classify(record)` : Method which takes a record which the method will classify. It should return this classification.**
- **`displayTree()` : Method which displays to screen the decision tree which was trained.**

You have flexibility within your code on how things are passed or displayed, but be sure to meet the design requirements. Private helper methods may be declared but should be declared and commented appropriately.

In [ ]:
```
#Write code here
```

5. Provide code that runs the training algorithm for training based on your provided data. Use the records from Question 3 to classify each record. Finally, output your decision tree.

```
In [ ]: #Write code here
```

Part 3: Analysis

6. Do your calculations for the Naive Bayes classifier result in the same classifier as the ID3 Decision Tree generates? If they differ, why do you think that is? If they are the same, do you think that they would always be? Explain your answer in detail.

Type answer here!