Lab 2 – Classification and Prediction #1

Data Mining, Spring 2018

Carol (berm@itu.dk) | Daniel (dafr@itu.dk) | Stefania (stsa@itu.dk)

IT UNIVERSITY OF COPENHAGEN

Feedback on last week's lab

Today's Lab: Classification and Prediction round 1

Mushrooms!

Classification & Prediction #1

- Today you will be working with data concerning mushrooms.

- Objective: build a classifier to predict if a given mushroom is edible or poisonous.

- You will build two classifiers:

- One using the ID3 algorithm

- The other using the kNN algorithm

- Compare their accuracy when classifying the mushroom data.

- Code is provided to help you load data and convert it to Java objects.

Plan of attack

- First take a look at the code provided.

- Then decide which of the two classifiers to begin with
  - kNN is the simplest of the two, and thus easier to implement.
    - Pg. 422-423 in book.
  - ID3 is more complicated and takes a bit longer getting started with
    - Pg. 332-340 in book.
    - Decision tree data structure needed that can be used in the algorithm but also for classification of test tuples.
    - Visualization of decision tree?
    - Where to split data into training and test data?

- After implementing each classifier compare the two's accuracy.

IT UNIVERSITY OF COPENHAGEN

The data

- Has been cleaned – no preprocessing needed.
    - No missing values
- 3000 tuples
- 22 attributes
    - Mixture of nominal and binary attributes.
- The mushroom data can be found in the agaricus-lepiotadata.txt file in the java-project
- An explanation found in the agaricus-lepiotaexplanation.txt file is also included
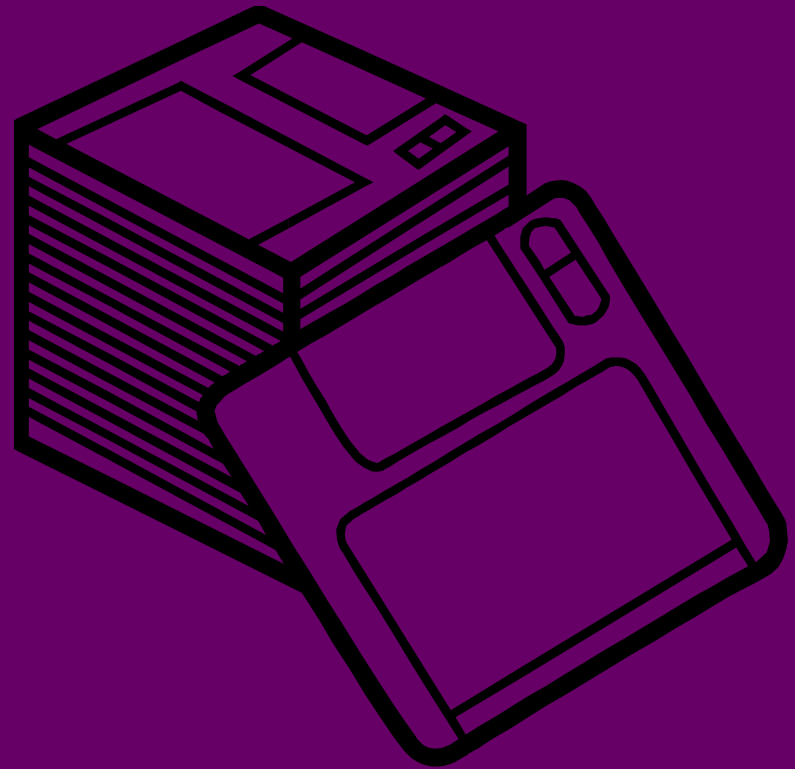
The code provided

- **Mushroom.java**
  - A mushroom class used to store data for each mushroom
  - Utilizes a lot of enums

- **CSVFileReader.java, DataManager.java**
  - Loads the data and creates the corresponding mushroom objects

- **Main.java**
  - Initializes the data loading process

Help working with enums

- Get all possible values for an enum:

```
for(Cap_Shape shap : Cap_Shape.values())
{
        System.out.println(shap.toString());
}
```

- Get a mushroom's enum value for a specific enum class:

```
mushroom.getAttributeValue(Cap_Shape.class);
```

- Get a list of all the attributes a mushroom has:

```
mushroom.getAttributeList();
```

Thanks for listening!