Artificial Intelligence I 2023/2024

Week 7 Tutorial and Additional Exercises

WEKA GUI & WEKA Jupyter Server

School of Computer Science

26th February 2024

In this tutorial...

In this tutorial we will be covering

- WEKA-GUI and WEKA in Jupyter Server

Before you begin

- School's Jupyter server can be accessed at `https://jupyterhub.oc1.aws.cs.bham.ac.uk`.
- When using school's Jupyter server, note that datasets used in this tutorial can only be uploaded to the folder `/jupyter/work`.
- First line in the notebook must locate the weka-jar file.
- Refer to Weka Manual Chapter 18 for more details.

Exercise 1: Applying Weka Filters in Jupyter Server

Implement the following steps in jupyter notebook.

- Load the weather.arff dataset. Use the filter `WEKA.filters.unsupervised.attribute.Remove` to remove the humidity attribute.
  - Hint 1: You can do it either via the `.setAttributeIndices(" ")` method of remove filter, or set the options via strings.
  - Remark 1: Make sure that the correct index of the attribute 'humidity' is used.

Exercise 2

Consider the weather.arff dataset. Suppose you want to classify the dataset into three classes based on temperature attribute. Try the following in both Weka GUI and Weka Jupyter Server.

- Directly set temperature as the class attribute. Is it possible to implement logistic regression on the new dataset?
- Now, use discretize filter to discretize the temperature attribute into three ranges. Set temperature as the class attribute. Report the classes. Run logistic regression and report its prediction accuracy on the training set.

Exercise 3: SimpleLinearRegression in WEKA

Simplelinear regression is used to describe the functional relationship between one feature attribute and a class attribute.
For the weather.arff dataset, find the simple linear regression model between the temperature attribute (feature) and humidity attribute (class). Do this in WEKA GUI and WEKA Jupyter server.
(Hint: You may want to pre-process data to have only one feature and class attribute. Which filters will you use?)

Exercise 4: Sentiment Analysis

Repeat the sentiment analysis exercise of Week 5 lecture in Jupyter server.
Describe the pre-processing tools used.

Exercise 5: Dominant Colour Extraction

Repeat the dominant colour extraction exercise of Week 5 lecture in Jupyter server.

- Run the K-means algorithm for $K = 4$ and $K = 8$ with kmeans++ as the initialization. Report and compare the WCSS when evaluated on training set.