Homework-1: Using Weka

Due Monday, September 9, 2019

30 points

Please write your answers to the Weka tutorial (which is Homework-1) on this page and turn it in. You only need to write answers where indicated, but you should think about the answers to the other questions. Be sure to complete the entire tutorial since it is important that you understand how to use Weka.

You may do this homework assignment jointly with another student (no more than 2 students working together). In that case, please submit just one answer sheet with both names printed at the top.

- Questions-1:

  A. How many instances are there in the data file? _____

  B. How many different values can gill-color take on? _____

  C. Which values of the gill-color attriute result in only edible mushrooms?

  _____

- Questions-2: Use a text editor to view the ARFF file representing the mushroom data.

  i. What does the first line in the file tell you? _____

  ii. What do the next 23 lines tell you?_____

  iii. What do the rest of the lines tell you? _____

- Questions-3:

  i. Draw the decision tree that was developed.

  ii. How many instances were classified correctly? _____
  How many instances were classified incorrectly? _____

- Questions-4:

i. What attribute is at the root of the decision tree?_____

ii. Consider the path cap-color=w, gill-spacing=c, population=a.
What is the class value assigned to instances that
follow this path?_____

iii. How many paths in the decision tree lead to a leaf node where some instances
are classified incorrectly? _____

iv. How many instances were classified correctly? _____

How many instances were classified incorrectly? _____

- Questions-5:

i. How many instances are incorrectly classified? _____

ii. What does the diagonal of the confusion matrix tell you?

iii. Which class value (p, or e) did the classifier get
wrong most often?_____

- Questions-6:

i. How many instances were used for training when there is a 50% split? _____
How many for testing? _____

ii. What was the percent success rate when there is a 50% split? _____

iii. How many instances were used for training when there is a 35% split? _____
How many for testing? _____

iv. What was the percent success rate when there is a 35% split? _____

v. How many instances were used for training when there is a 5% split?_____
How many for testing?_____

vi. What is the percent success rate when there is a 5% split? _____

vii. What is the error rate under each of the different splits?

50% split: _____

35% split: _____

5% split: _____

viii. What is causing the differences in classification error rate under the different splits?

- Questions-7:

i. How many instances from the test set were classified correctly?_____

ii. Draw the confusion matrix.