Practical 7

**Classification Review: 1R, Naïve Bayes, Decision Trees**

# What are we doing?

Using what you learned from lectures (Lecture 5, 6 & 7) and relevant reading materials, you will answer some review questions. These questions are for your self-review on topics covered: Supervised learning (classification) and basic classification methods: 1R, Naïve Bayes and Decision Trees.

You will need to review lecture and reading materials or seek for other resources (e.g. Googling), in order to answer questions.

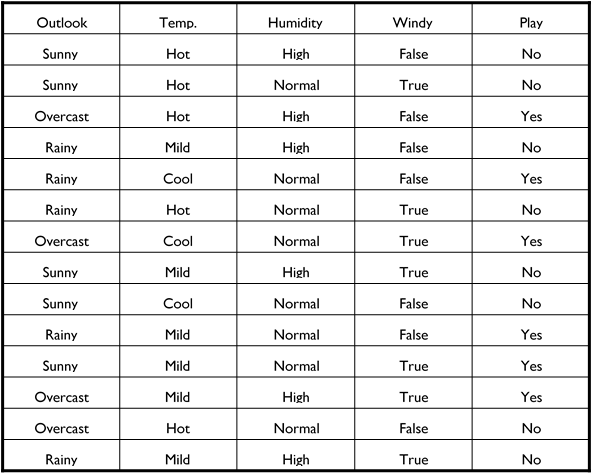
**Submission:**

You are required to submit one document containing your answers via the weekly-practical submission box (available on CP1407 LearnJCU)

For Laboratory questions, screen capture your computer screen after the completion of each task and include the captured image in your document to submit.

# Review Questions

1. Compute the 1-R classification on the following dataset. Based on the classification result, derive a contingency table and compute sensitivity, specificity, recall, precision, FP-rate and TP-rate.



1. Suppose we have data on a few individuals randomly surveyed. The following table gives the data on these individuals as to their response showing interest to promotional offers made in areas of Finance/Investment, Travel/Tour, Reading/Magazines, and Health/Diet. For this example, we use Sex as the output attribute whose value is to be predicted.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Finance/Investment Promotion | Travel/Tour Promotion | Reading/Magazine Promotion | Health/Diet Promotion | Sex |
| Yes | No | Yes | No | Male |
| Yes | Yes | No | No | Male |
| No | Yes | Yes | Yes | Female |
| No | Yes | No | Yes | Male |
| Yes | Yes | Yes | Yes | Female |
| No | No | Yes | No | Female |
| Yes | No | No | No | Male |
| Yes | Yes | No | No | Male |
| No | No | No | Yes | Female |
| Yes | No | No | No | Male |

Use the data together with the Naïve Bayes classifier to perform a new classification for the following new instance:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Finance/Investment Promotion | Travel/Tour Promotion | Reading/Magazine Promotion | Health/Diet Promotion | Sex |
| **No** | **Yes** | **Yes** | **No** | **?** |

1. List and justify major strengths and weaknesses of Naïve Bayes’ Algorithm.
2. List and justify major strengths and weaknesses of decision tree algorithms.

# Laboratory Questions

The data set in the table below contains data about heart disease and its conditions. The class label P means that heart disease is present and the class label N that the disease is absent.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Body  Weight | Body Height | Blood Pressure | Blood Sugar Level | Habit | Class |
| Heavy | Short | High | 3 | Smoker | P |
| Heavy | Short | High | 1 | Non-smoker | P |
| Normal | Tall | Normal | 3 | Non-smoker | N |
| Heavy | Tall | Normal | 2 | Smoker | N |
| Low | Medium | Normal | 2 | Non-smoker | N |
| Low | Tall | Normal | 1 | Non-smoker | P |
| Normal | Medium | High | 3 | Smoker | P |
| Low | Short | High | 2 | Smoker | P |
| Heavy | Tall | High | 2 | Non-smoker | P |
| Low | Medium | Normal | 3 | Smoker | P |
| Heavy | medium | Normal | 3 | Non-smoker | N |

1. Use the dataset above as a training set and perform the following tasks in WEKA:
   1. Use the J48 algorithm to construct a decision tree under each of the test options. Use the validation data set in the table shown below as a test set for the Supplied test set option.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Body  Weight | Body Height | Blood Pressure | Blood Sugar Level | Habit | Class |
| Heavy | Short | High | 2 | Smoker | P |
| Heavy | Tall | Normal | 1 | Smoker | N |
| Heavy | Medium | Normal | 3 | Smoker | N |
| Low | Short | Normal | 3 | Smoker | N |
| Low | Medium | High | 1 | Non-smoker | N |
| Low | Medium | High | 3 | Non-smoker | P |

* 1. Use the decision tree obtained from the previous task ( task a) ) to determine the classes for the unseen data records as below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Body  Weight | Body Height | Blood Pressure | Blood Sugar Level | Habit | Class |
| Heavy | Medium | High | 1 | Smoker |  |
| Heavy | Medium | Normal | 3 | Smoker |  |

1. Perform decision tree induction in WEKA using different decision tree induction algorithms (RandomTree, J48, REPTree). Compare the resulting trees and the measures of accuracy. Switch the pruning parameter on and off to observe the unpruned and pruned trees.