

COMP3620 Assignment 4: Learning

1 Task 1 (20 points)

Please download the two files

```
http://rsise.anu.edu.au/~kee/COMP3620/training.data
http://rsise.anu.edu.au/~kee/COMP3620/test.data.
```

The file training.data contains a collection of pairs (x, y) , where x is a list of characters and y its label (positive or negative). These data are generated by an underlying process P unknown to you. The file test.data contains a collection of pairs $(x, -)$ generated from the same process P , but with the labels withdrawn. The task is to predict the missing labels in test.data.

The following two items are required for this assignment.

1. A file called predictions.txt containing your prediction of the label of each x in test.data.
2. A report (in PDF format) detailing the process/algorithm used to produce the predictions.

The file predictions.txt should contain just a list of labels, the n -th label being the prediction for the n -th entry in test.data.

This problem is open-ended by design. There is no one unique right solution. There are, however, objective measures by which one can judge the quality of a solution. Submissions will be judged on the following two measures.

The first measure is quantitative, being the accuracy of the predictions on the test data. A maximum of 8 points is available for this part. So, for example, a submission that gets 80% of the labels right will be awarded $0.8 \times 8 = 6.4$ points. (Marks may be rounded.)

The second measure is qualitative. The design and execution of the learning algorithm/process and the rationale behind major decisions will be judged based on established machine learning principles discussed in the lectures. A maximum of 12 points is available for this part. In addition to a description of the submitted solution, you may wish to discuss the failed experiments/ideas you tried on the way to arriving at the final solution.

You're free to use any publicly available resources to solve this problem.

2 Task 2 (5 points)

This problem is for COMP6320 students only.

Design and execute an experiment to either confirm or discredit the solution to the Monty Hall problem obtained via Bayes Rule. The report describing the solution should provide

1. the rationale behind the experiment design; and
2. experimental results obtained and conclusions drawn.

3 Submission Instructions

Please submit a single zip file containing all relevant files. To submit the file (for example assign.zip)

- Log on to partch.anu.edu.au
- In the home directory type
`submit comp3620 lab4 assign.zip`