Coursework FAQs

EMATM0044 2020

1 Coursework Part 1

General and formatting

1. Q: How should I submit the coursework?

   A: Please submit the coursework in two parts:

   1) A pdf file containing written answers, plots, and figures.

   2) A zip file containing your code. Your code should be either .py files or .ipynb files.

   Please do not put code in the written report section.

2. Q: How should I format the written report?

   A: Please use LaTeXor MS Word. If you use Word, please convert the files to a pdf before submitting.

3. Q: Should I combine the written answers for Q1, Q2, Q3 into one document?

   A: Yes, please combine the written answers for Q1, Q2, Q3 into one document. Code files can be separate documents

1.1 Q1

1. Q: Can I use built in models from scikit-learn?

   A: Yes, you can use models from scikit-learn or any other library

2. Q: Do I need to code my own model for Q1?

   A: No, you do not need to

3. Q: What is a baseline model?

   A: A baseline model is a simple model than you can compare your model against

4. Q: What is a hyperparameter?

   A: A hyperparameter is a value, such as k in k-nearest neighbours or maximum depth for a decision tree, that can alter the performance of your algorithm. You should use an appropriate method to select a good value of these parameters. Worksheet 13, the section on model selection, will be useful here.

5. Q: What if my model has no hyperparameters?

   A: That's ok, but please discuss this aspect of your model in the report.

6. Q: Should I choose two different types of algorithm (e.g. clustering/regression) or two versions of the same type (e.g. two different regression algorithms)

   A: You should choose two versions of the same type.

1.2 Q2

1. Q: Can I use the built in scikit-learn library?

   A: No. This question assesses your ability to code this model from scratch.

2. Q: Do I need to evaluate the performance of the decision tree?

   A: No

3. Q: Do I need to compare the performance of my decision tree against the model from scikit-learn?

   A: No

4. Q: I have been able to find some useful online references for approaches others have taken towards building a decision tree from scratch. What would be an appropriate referencing convention to use?

   A: The following give some good guidance `https://student.unsw.edu.au/how-do-i-cite-electronic-sources`

5. Q: When we bin the data into high, medium, and low, some datapoints have the same values but different classifications. How should we deal with this?

   A: This means that the data is noisy, and there will not be a perfect classification.

1.3 Q3

1. Q: If a question is not applicable to this dataset, can I say that it is not applicable?

   A: Yes, and please briefly explain why not.

2. Q: Would you expect a Datasheet layout similar to the examples in the paper appendix?

   A: No, plain formatting is fine

3. Q: In the circumstance that you are paid by an organisation to carry out research, does that mean that the research project is in fact funded by the organisation that employs you?

   A: The source of funding for a research project is usually in terms of grants, so if an organisation provided a grant to fund the research this would mean that the research is funded by them.

2 Coursework Part 2

General and formatting

1. Q: Do I need to submit code?

   A: No, you do not need to submit any code

2. Q: Can I submit code as part of my answers?

   A: For this part of the coursework, I would expect to see the equations written out, although it is fine to use python or other software to do the actual calculations

3. Q: Can I handwrite my answers and scan them in?

   A: Please try to prepare as much of the document as possible using LaTeXor Word. It is possible to typeset equations in Word, please let me know if you need help with this. If needed, you can handwrite diagrams and equations and insert them into your document.

2.1 Q1

1. Q: Why is the uniform weighting relevant?

   A: You should think of each nearest neighbour as contributing equally to the classification

2.2 Q2

1. Q: Should we use natural log or another logarithm?

   A: You should use natural log

2.3 Q3

1. Q: What is the identity matrix?

   A: The identity matrix I is a matrix with 1 along the diagonal and 0 everywhere else. Then for any matrix M, $I \cdot M = M = M \cdot I$

2.4 Q4

1. Q: Should there be arrows on the graph?

   A: The graph is undirected (there are no arrows). However, for breadth-first and depth-first search please use a list of visited nodes so that you do not run into cycles.

2. Q: How should we interpret negative costs?

   A: You can think of negative costs as part of a journey that you want to make to e.g. see a particular landmark

2.5 Q5

1. Q: What does it mean for the space to move?

   A: You can think of the space moving , rather than the pieces moving. This makes the puzzle easier to model. So if the space moves up, the piece above it moves down.

2. Q: Do you still follow the move order up, left, down, right in A* search?

   A: Yes, but the heuristic values may override this order.

2.6 Q6

2.7 Q7