PS3192 Portfolio Coursework: FAQ and Clarifications

Matteo Lisi

Table of contents

General Guidance

How should I structure my answers?

There is no rigid template to follow. Use a structure that helps communicate your analysis clearly. Subheadings are optional but can help organise your work, especially for longer answers (e.g., Question 2).

Is there a word limit per question?

The total portfolio should be no more than **5,000 words across three questions**. This is an upper limit rather than a target. For example, Question 2 can typically be answered in **600–800 words**. Some questions may require more words depending on the depth of analysis or inclusion of examples, while others may be shorter if your analysis is clearly conveyed through plots and captions.

Is there a lower word limit?

There is no strict lower word limit. Some answers may naturally be shorter than others, especially if you use clear visualizations, concise explanations, or code outputs to communicate your analysis.

As a rough guide, around 500 words per question is usually enough to cover the key points clearly for questions 3-4. Some answers might be longer, especially if you discuss multiple models or methods.

Focus on clarity, completeness, and quality rather than aiming for a specific word count.

Do figures need captions? What about explanatory text in the main body?

If the caption provides a complete explanation of the plot (i.e., what is shown, patterns or trends, and interpretation), then that may be sufficient. However, some questions (e.g., Q1) also expect a brief discussion of the figure in the main text. Use your judgment—the goal is clarity.

Question 1: Data Visualisation

Can I use colour in the plots? Do they have to follow APA style?

Yes, **use colour**! You are not required to follow APA-style black-and-white formatting. Use colour to make your plots clearer and more readable. Try to use accessible colour palettes (e.g., colorblind-friendly schemes).

What counts as a STEM subject?

You are free to define STEM in a way that is either broad or strict. A strict definition would typically include:

- Engineering

- Physical Sciences

- Biology and Life Sciences

- Agriculture and Natural Resources

The inclusion of Health-related fields is more ambiguous. Some subfields (e.g., "Nutrition Sciences") may be considered STEM by some definitions, but others (e.g., "Health and Medical Administrative Services") are less clearly STEM. Justify your decision and apply your definition consistently.

Question 2: Machine Learning Concepts

How many words should I write?

Around **600–800 words** should be sufficient. This is a guide rather than a requirement. You can write more if you're including detailed examples or discussing additional literature. Examples are optional but can strengthen your answer.

Do I need an introduction and conclusion?

You do not need explicitly labeled sections, but a clear structure is helpful. Use subheadings if they help organise your argument.

Should I discuss both k-fold and leave-one-out cross-validation (LOOCV)?

Both are forms of cross-validation. In **k-fold CV**, the data is split into k subsets; each subset takes a turn as the test set. In **LOOCV**, k equals the number of observations.

You are not required to describe both methods in detail, but you are welcome to mention both if it helps explain the purpose and use of cross-validation.

Question 3: Supervised Learning

What structure should I use in my answer?

A structured main text is appropriate and encouraged. For example, a possible structure is:

1. **Introduction**
   State your chosen dataset, describe the outcome variable and available features.

2. **Modelling Approach**
   Describe the model used (e.g., decision tree, logistic regression), any preprocessing steps, and model selection choices (e.g., tree pruning).

3. **Results**
   Report model outputs (e.g., tree size, used features), include relevant plots, and report confusion matrix or other metrics.

4. **Discussion**
   Interpret key findings, discuss which variables were most predictive, potential overfitting, performance comparisons, and how model performance could be improved.

Optional but Recommended: RMarkdown

It is optional but recommended that you use **RMarkdown** to write your portfolio. This allows you to:

- Integrate plots, narrative, and code in one document

- Easily export to Word format

- Include both R and MATLAB code in a clean and readable way

To include MATLAB code (non-executable) in RMarkdown start a code chunk as follow:

```
```{matlab}
% This is MATLAB code
% It is not executed but displayed neatly
2 + 2
```
```

This will render to:

```
% This is MATLAB code
% It is not executed but displayed neatly
2 + 2
```

Code, Output, and Comments

Can I show output (e.g., plots, summaries) in the main text without including the code?

Yes — this is a recommended approach. Your main text should focus on results and interpretation, and you can include figures or summary tables without showing the code that produced them. Just make sure to **describe what the output shows**, especially if it's not self-explanatory.

Do I need to include all code output?

No. You only need to include **key outputs** that are relevant to your analysis (e.g., model summaries, key plots). For data exploration or setup code (e.g., checking variable types), you can summarise what you did and found in words without showing the output.

Can I keep # comments in my R code?

Yes — please do! Comments in code are encouraged and part of the **code readability** marking criteria. They help the reader understand what your code does and why. Use comments to explain the purpose of key sections or clarify any non-obvious steps.

Do I have to use a specific structure to separate code and output?

There's no rigid format, but a good approach is:

- **Main text:** Focus on narrative, explanation, and key outputs. Use `echo=FALSE` to hide the code but show the results.
- **Appendix:** Include all code, ideally with `eval=FALSE` so it is shown but not re-executed.

This keeps the main body clean while ensuring reproducibility.

Final notes

- Always include *all* your code in the appendix.
- The appendix does not count toward your word limit.
- You may use more than one model or approach in Questions 3, 4, or 5, especially if comparing performance.

If you have further questions, don't hesitate to get in touch. Good luck!