Coursework

Question 1 (40 pts)

In this coursework you should create an artificial dataset and test two classification algorithms on the data you have produced. The dataset should be produced by picking a number of dimensions, for example $d = 5$. Now select $c = 3$ 'center points' in that five dimensional space, in the first instance pick each component uniformly in the interval $[-1, 1]$. Now for each center point, make $n = 20$ datapoints from a multivariate normal distribution centered on the center point with variance σ^2, with $\sigma = 0.5$.

Now run two different classification algorithms on these data and assess their accuracy. They should be pretty accurate. Consider three different ways to change the simulated data to reduce their accuracy, this could include changing d, σ, n, for high marks you should include at least more complicated manipulation, such as the balance in the number of points for each center point or changing the shape of the cloud of simulated data, for full marks this might include cloud shapes that are nothing like a ball, even the ball used in Australian or American football.

Produce plots showing how the accuracy changes under these manipulations and try to use your investigation to comment on the relative advantages and disadvantages of the two algorithms you are looking at. In looking at accuracy consider different measures of accuracy and the accuracy of a baseline model. You should make sure any assessment is not restricted to the data used in train models or decide on hyper-parameters: it is important to hold aside testing data.

In your report you should explain your decisions. You code will not be marked for elegance, but it should run correctly; it is expected you will use Python, but any of Python, Julia or R is fine. Do not include screenshots of graphs, they should be imported directly; resize them to the correct size before importing them, if the labels are tiny the graphs will not be marked. Make sure figure captions are descriptive, it is better to have some overlap between figure captions and the main text than to have figure captions that are not reasonably self-contained.

As a rough guide to marking:

- Initial production of simulated data, including some graphs or other approaches to visualisation. (6 marks).

- Two algorithms should be tested, if only one algorithm is included the 28 available marks will be halved.

- Overall presentation (3 marks), including use of appropriate sections, plots, diagrams, or tables to make your point. Do not include code snippets in the report. Instead, describe in words or equations what you are implementing. Format equations correctly.

- Suitable choice of algorithms (2 marks).

- Suitable choice of evaluation for algorithms (4 marks).

- Two more obvious manipulations and an illustration of who they change claissification accuracy, with a good and sensible set of measurements of accuracy. (4 marks)

- A more advance manipulation of the simulated data along with the corresponding information on how this affects classification. (4 marks)

- Bonus marks if the advanced manipulation is very sophisticated. (2 marks)

- Insight into what this experiments tells us about the two classification algorithm. (3 marks)

- There are some marks (6 marks) for something surprising and unusual. This is to allow us to give credit for exceptional work, but you should not expend too much time chasing these final marks, concentrate on the standard requirements to get a good mark.

Question 2 (10 marks)

New advances in AI will cause substantial changes in higher education, in what we teach, in how we teach, how we examine. These changes come with complex ethical challenges, for example, in how we assess students in a fair way while teaching them to make use of AI tools. It is already clear that students are using AI to do coursework questions, particularly those with essay type answers; this makes the marks or degree-grades less meaningful, narrowing the opportunity for particularly able or hard-working students to succeed in competition with students from more privileged backgrounds.

Indeed, more generally, AI may advantage some students over other, or may amplify existing biases; more optimistically, AI may help address ethical challenges already embedded in higher education. Discuss this.

Report

Your report should be no longer than five pages, excluding any references. It is expected that Question 2 would occupy about a fifth of this space; use an 11 or 12pt font and do not try tricks like expanding the margin to fit in more text, shorter is better than longer.

Your report must be submitted in pdf and should be prepared in LaTeX; overleaf is a good approach, but not required as long as LaTeX has been used[1]. As always when using LaTeX, give yourself over to defaults, our expectation of what a document should look like has been conditioned on LaTeX, so it is best not to try to override the look of the document. I have included a template but you need not use that.

Avoid code snippets in the report unless that feels like the best way to illustrate some subtle aspect of an algorithm; do always though consider a mathematical description if possible. You will be asked to submit code and it may be tested to make sure it works and matches your report. It will not, however, be marked in and of itself.

Submission

The deadline for report and code: 13h00 (GMT+1) on XXX, there will be a submission point on Blackboard under the "assessment, submission and feedback" link. Please upload the following two files:

1. Your report as a PDF with filename ¡student_number¿.pdf, where "¡student_number" is replaced by your student number, not your username. Upload this to the submission point "Introduction to AI Coursework (Turnitin)".

2. Your code inside a single zip file with filename ¡student_number¿.zip. Inside the zip file there should be a single folder containing your code, with your student number as the folder name. Please remove datasets

---

[1] R-markdown and some other notebook-based environments typeset using LaTeX, this is acceptable

and other large files to minimise the upload size - we only need the code itself. Upload this file to the submission point "Code for Introduction to AI Coursework".

We may review your Python code by eye but your marks will be based on the contents of your report, with the code used to check how you carried out the experiments described in your report. We will not give marks for the coding style, comments, or organisation of the code. Code written in Julia or R is also acceptable as is the use of a standard notebook format. If you are particularly keen on another programming language let me know and I will consider this; I would accept other modern languages such as Rust, but outmoded or unsuitable languages like C++, Java or MATLAB would not be allowed.

Please do not include your name in the report text itself: to ensure fairness, we mark the reports anonymously.

Avoiding Academic Offences: Please re-read the university's plagiarism rules to make sure you do not break any rules. Academic offences include submission of work that is not your own, falsification of data / evidence or the use of materials without appropriate referencing. Note that sharing your report with others is also not allowed. These offences are all taken very seriously by the University and we have very little leeway within the framework the University has set out. Do not copy text directly from your sources - always rewrite in your own words and provide a citation. Work independently – do not share your code or reports with others; you can, of course, discuss your work with your classmates, but do not share text or code.

Suspected offences will be dealt with in accordance with the University's policies and procedures. If an academic offence is suspected in your work, you will be asked to attend an interview with senior members of the school, where you will be given the opportunity to defend your work. The plagiarism panel can apply a range of penalties, depending on the severity of the offence. These include a requirement to resubmit work, capping of grades and the award of no mark for an element of assessment. Again, we are not in a position to be lenient here, the academic offences procedure is not one we control.

Extensions and Exceptional Circumstances

If the completion of your assignment has been significantly disrupted by serious health conditions or personal problems, or other serious issues, you can apply for consideration in accordance with the normal university policy and processes. Students should refer to the guidance and complete the application forms as soon as possible when the problem occurs. Please see the guidance below and discuss with your personal tutor for more advice:

- `www.bristol.ac.uk/students/support/academic-advice/`
  `assessment-support/request-a-coursework-extension/`

- `www.bristol.ac.uk/students/support/academic-advice/`
  `assessment-support/exceptional-circumstances/`