61011 short answer exam report

Q1 (out of 7 marks, mean: 4.2 s.d.: 2.0)

(a) Candidates generally performed well on this part of the question. Partial credit was given for …

(b) This was book work. A number of students did not state the convergence theorem, but simply wrote pseudocode to implement a perceptron classifier. Full marks were given in instances where the candidate was not completely precise (e.g. omitting mention of perfect classification)

(c) Part c challenged the candidates to understand how a logistic regression would converge when perfect classification was not possible. A small proportion of students incorrectly stated that the algorithm would not converge.

Q2 (out of 8 marks, mean: 6.3 s.d.: 1.6)

(a) Candidate were awarded two marks for any two sensible statements comparing the decision boundaries of linear SVMs and perceptrons. This included any mention of slack variables (omitted on the original mark scheme). Marks were not awarded for any discussion about kernels or projection to higher dimensions, as the question concerned linear SVMs only

(b) Book work was completed well by the vast majority of candidates, as reflected by the high means and low standard deviation.

Q3 (out of 6 marks, mean: 3.8 s.d.:1.5)

(a) Candidates were asked to calculate sensitivity, specificity and F1 score from a given confusion matrix. One common mistake was using the wrong definition, reporting positive and negative predictive value instead

(b) Most correctly identified the best classifier. A mark was given for correct identification and for a sensible justification. This justification required the candidate to mention that the classifier had the highest true positive AND lowest false negative rate. A further mark was given for identifying the random classifier as the worst performing model. This proved tricky for many, who did not realise that classifiers in the bottom half of a ROC curve are more informative than a random classifer.

Q4 (out of 4 marks, mean:1.9 s.d.:1.2)

Question 4 had a typo that was not discovered during exam preparation or moderation. Our approach to mitigate the impact of this is discussed below.

Ignoring this, question 4 was particularly challenging, involving events (test result) that were conditioned on two prior events (having a disease, having test 1 or test 2). Most students failed to take into account one of these, leading to an incorrect final answer of ~45.5% (for which partial credit was

given). Application of Bayes rule was not straightforward in this question, and we were surprised to see very few candidates attempt to solve the question using probability trees.

MCQs

The average mark was 16/25 (s.d 3.5). The minimum and maximum marks were 6 and 25, respectively

Mark moderation

To minimize the impact of the typos in question 4, we considered two strategies: (a) grade using the question as it stands, (b) regrade excluding the question all together (i.e. marked out of 21). To avoid penalizing candidates who performed well on question 4, we finally decided on a hybrid approach, in which the final mark was the higher grade of the two strategies.

We checked all 130 students, to see if the strategy caused any to move either up or down a degree boundary. No student was disadvantaged. Furthermore, no student moved up over the 50 or 60 boundary, but 5 students moved over the distinction line to just above 70. The average grade raises from 67% (s.d. 7.5%) to 69.4% (s.d. 7.8%).