

PGT Exam Performance Feedback 2016/2017 Semester 1

COMP61011 Foundations of Machine Learning

Ross King
Xiao-Jun Zeng

Comments Please see the attached report.

Comp 61011 Exam Feedbacks

Summary:

- Among 58 students who took the exam, the average mark is 60.1%.
- 17 or 31% students received a score of 70% or better.
- 7 or 12.1% students received a score between 60% and 69%.
- 16 or 27.6% students received a score between 50% and 59%.
- 15 or 25.9% students received a score between 40% and 49%.
- 2 or 3.4% students received a score below 40%.

General Feedback for Question A:

- Among 58 students who took the exam, the average mark for Part A is 53%.
- 14 or 24.1% students received a score of 70% or better (i.e., 10.5 marks or more out of the total 15 marks)
- 10 or 17.2% students received a score between 60% and 69%, (i.e., 9 or 10 marks).
- 7 or 12.1% students received a score between 50% and 59%, (i.e., 7.5 or 8 marks)
- 10 or 17.2% students received a score between 40% and 49%, (i.e., 6 or 7 marks)
- 17 or 29.3% students received a score below 40% (i.e., less than 6 marks)

In general the students exhibited good working knowledge of the key concepts examined: Bayes theorem, and random forests (an decision tree based ensemble method).

Detailed Feedbacks for Question A:

The students struggled with the more advanced questions 2) iv, v, vi which examined concepts related to objective/subjective interpretations of probabilities, and decision theory.

General Feedback for Question B:

- Among 58 students who took the exam, the average mark for Part B is 67%.
- 26 or 44.8% students received a score of 70% or better (i.e., 11 marks or more out of the total 15 marks)
- 14 or 24.1% students received a score between 60% and 69%, (i.e., 9 or 10 marks).
- 6 or 10.3% students received a score between 50% and 59%, (i.e., 8 marks)
- 12 or 20.7% students received a score between 40% and 49%, (i.e., 6 or 7 marks)
- No student received a score lower than 40% (i.e., less than 6 marks)

General speaking, the students' performance in Part B is very satisfactory with the average mark as 67%, 45% students with a distinction score, and no students with less than 40% score. However, it is still disappointed to see that 20.7% students failed to get the pass score 50%.

Detailed Feedbacks for Question B:

- Questions 1 and 2. Most students answered this question correctly.
- Question 3. The question is “You are applying SVM with the following error function to classify a dataset $E = \sum_{i=1}^N \max\{0, 1 - y_i f(x_i)\} + \frac{1}{2} \sum_{j=1}^d w_j^2$. After a few tries, you realise that there are some “outlier” datapoints. Now answer the following questions: i) How do you modify the above error function to handle these “outlier” datapoints? ii) What is the meaning of each new item in your modified error function above? (2 Mark)”.

More than 1/3 students made some mistake with this question. The correct answer to the modified error function is

$$E = \sum_{i=1}^N \max\{0, 1 - y_i f(x_i) - \xi_i\} + \frac{1}{2} \sum_{j=1}^d w_j^2 + c \sum_{i=1}^N \xi_i$$

The most common mistake is the incorrect error function as below

$$E = \sum_{i=1}^N \max\{0, 1 - y_i f(x_i)\} + \frac{1}{2} \sum_{j=1}^d w_j^2 + c \sum_{i=1}^N \xi_i$$

That is, ξ_i is missed. However, if such a missing ξ_i makes impossible to handle the “outlier” and so is a mistake. Further some students did not include the sum at the end of the modified error function.

- Question 4. Most students answered this question correctly.
- Question 5. The question asked is to calculate the entropy and information gain, and then based the calculation result to decide how to split a decision tree. About half of the students knew how to answer the question, but 1/5 students did not know how to answer this question at all. For the rest students who knew how to answer this questions in different degrees, the common mistakes are that the required formula were not given, the used formulas are incorrect, some calculation steps are missed, and the calculation results are incorrect.

General Feedback for Question C (MCQ):

- Among 58 students who took the exam, the average mark for Part B is 60%.
- 12 or 20.7% students received a score of 70% or better (i.e., 14 marks or more out of the total 20 marks)
- 19 or 32.6% students received a score between 60% and 69%, (i.e., 12 or 13 marks).
- 19 or 32.6% students received a score between 50% and 59%, (i.e., 10 or 11 marks)
- 8 or 13.9% students received a score below 50%, (i.e., 9 or less marks)

General speaking, the students' performance in Part C is not as satisfactory as we had hoped, giving the average score as 60%.

Detailed Feedbacks for Question C:

- The most difficult questions are questions 10 and 20, and each of them was answered incorrectly by more than 70% students.
- The other difficult questions are questions 9, 15, 18 and 19 and each of them was answered incorrectly by more than 60% students.