

**Comments** Most people could answer q1a and 1b correctly, about the perceptron and KNN models. On q1c, several people made simple errors of calculation or writing of the entropy equation - being careful is important here. I docked 1 mark in this situation. On q1d, many people fell into the trap of not reading the question properly. The question clearly asked you to name a supervised learning model and an associated "model selection parameter". In this situation several people simply wrote everything they knew about perceptrons, everything they knew about decision trees, etc, without actually answering the question "name a model selection parameter". On q1e, only about half the class could correctly and fully describe the RF algorithm.

Regarding Question C.1, students did not perform well in general. For C.1(a), it is a question on book knowledge but requires the understanding on general clustering analysis. Most of students can give details of K-mean algorithm but did not even think about an essential step required by any clustering algorithms including the one in this question, i.e., selecting a specific distance metric before starting a clustering algorithm. For C.2(b), again, an answer to this question only needs book knowledge explicitly appearing in the lecture note but nearly 50% students failed to answer this straightforward question correctly. For C.2(c), it is regarding a typical application of clustering analysis, which requires the deep understanding in clustering analysis and problem solving skills by flexibly applying what has been learned to this problem. While only quite few students take the main aspects concerned by this real world problem into account, most of students simply gave a clustering algorithm without analysis and reasonable justification in their answers, which is insufficient and inevitably leads to a low mark.

---