Marking Correlation

Romain Lhotte & Paul Dubois*

April 13, 2023

Abstract

The evaluation of students' performance in higher education plays a crucial role in assessing the effectiveness of teaching methods and improving learning outcomes. In this study, we investigate the correlation between two teachers' grading practices in a deep learning course at the master's level, offered at CentraleSupélec. The two teachers, who have distinct teaching styles and backgrounds, were responsible for marking the final project oral presentation. Our results indicate a significant positive correlation between the two teachers' grading practices, suggesting that their assessments of students' performance are consistent. Our findings provide insight into the reliability and consistency of grading practices in higher education and can inform efforts to improve the quality of teaching and learning outcomes.

1 Introduction

In recent years, there has been a surge of interest in the field of deep learning, with applications ranging from computer vision and natural language processing to bio-informatics and medical diagnosis. As a result, there is an increasing demand for high-quality education and training programs in deep learning. In response to this demand, many universities and engineering schools have started offering courses and programs in deep learning at various levels, including undergraduate and graduate levels.

This study focuses on the evaluation of a deep learning course offered to third-year engineering students at CentraleSupélec (Paris, France), who are pursuing the "Vivant, Santé, Environnement" (VSE) path. The course was designed to provide students with a comprehensive understanding of the fundamental concepts, theories, and applications of deep learning. The course was structured into 10 teaching sessions, each spanning three hours. Each session comprised one hour of theoretical instruction followed by two hours of practical exercises. In addition to the classroom instruction, five Kaggle challenges were assigned to the students, who were expected to work on them independently and outside of class time. The final component of the course consisted of a group project, which required the 28 students to form 10 groups of 2-4 individuals. The project served as a comprehensive assessment of the students' proficiency in deep learning and required the application of the concepts and techniques covered in the course. Students were expected to present their project findings orally and respond to questions from the instructors.

The evaluation of students' performance in the course is a critical aspect of assessing the effectiveness of the teaching methods and the quality of the learning outcomes. In this study, we investigate the correlation between the grading practices of two instructors who taught the course, and the Kaggle assessments. Our goal is to evaluate the reliability and consistency of grading practices in higher education and to provide insights that can inform efforts to improve the quality of teaching and learning outcomes.

The findings of this study have implications for educators, administrators, and policymakers who are interested in improving the quality of education in deep learning and related fields. The results can also be used to inform the development of more effective evaluation frameworks and grading practices in higher education. Overall, this study contributes to the growing body of knowledge on the pedagogy of deep learning and its applications in various fields.

2 Methods

2.1 Participants

The participants in this study were 28 students enrolled in a deep learning course at the master's level offered at CentraleSupélec during the academic year 2022/2023. The class was taught by two instructors,

---

*both authors contributed equally to this work

both of whom had distinct teaching styles and backgrounds.

2.2 Evaluation Framework

The evaluation of students' performance in the course was based on six components, which accounted for the final grade.

The first five components were Kaggle challenges, each accounting for 6% of the total mark, adding up to 30% of the final mark. For each challenge, a marking scheme was established, which was publicly shared with the students:

3/6 achieved a better score than one typically obtained after a **basic** attempt of the challenge

4/6 achieved a better score than one typically obtained after a **fair** attempt of the challenge

5/6 achieved a better score than one typically obtained after an **advanced** attempt of the challenge

6/6 given to the **top 10** students of the class; to obtain 6/6, students also needed to meet requirement for 5/6 (the class is 28 students, so this score is reachable for a third of the class)

0/6 no attempt to the challenge or score *below* what a typical *basic* attempt of the challenge

Details on the exact scores level for each Kaggle challenges are explained in the supplementary material.

The final component was a project, which accounted for the remaining 70% of the final mark. Students completed the project in groups of 2-4, and the project was assessed based on an oral presentation of approximately 10 minutes, followed by approximately 10 minutes of questions. The final mark for the project was the average of the two independent marks given by the two instructors, who graded the projects separately.

2.3 Ethical Considerations

This study was conducted in accordance with the ethical guidelines of CentraleSupélec, and all participants were were anonymized for confidentiality purposes.

2.4 Normalizing

The Kaggles were originally graded out of 6, the project out of 20. We normalized all graded to be out of 100 (%), to ba able to compare mean and standard deviation.

3 Results

We rounded all numerical values to 10^-2 for readability purposes.

3.1 Dependence of assessments

Suppose that the score of students on each kaggle were independent. Then probability theory tells us that the standard deviation of the average kaggle score should be 9.85 (using $\text{Var}\left(\frac{\sum_{i=1}^{5} K_i}{5}\right) = \frac{\sum_{i=1}^{5} \text{Var}(K_i)}{5^2}$).

3.2 Distribution of the marks

3.3 Quantification of the inter assessments correlation

4 Discussion

5 Conclusion