DD2424: Assessment of the group project

Josephine Sullivan

1   Evaluation of deep learning knowledge acquired (20%)

After completing the project we would like each student to have acquired some specific knowledge in deep learning theory, algorithms and/or software packages that is tied to practical experience. Examples of such specific knowledge or skills would be:

- familiarity and ability to write network architectures and train them and run experiments with a recognized deep learning software package such as TensorFlow or PyTorch.

- familiarity with a common/popular network architecture, the pros/cons of this architecture and what it is commonly used for.

- familiarity with a popular training algorithm and understanding of how it works.

To grade this we would like each student to write down (on a separate page as an appendix in their project report and then refer to it in their presentation):

1. The **three main and separate deep learning skills/knowledge** they have acquired during the project and

2. A very brief description of how they acquired this skill/knowledge and the evidence that they acquired this skill/knowledge.

Then the grader will assess whether the project as a whole (oral presentation, report and ability to answer questions) reflects the student's assessment. The assessment will be done as follows. For each *skill* the grader will assign an assessment of 3, 2, or 1 where the numerical grades correspond to:

Score	Meaning	Awarded when
3	Agree with student's assessment	Student can talk/write knowledgeably and cogently about the skill and related issues and answer questions sensibly. Gives plenty of evidence of the tasks they performed to acquire this skill.
2	Partially agree with student's assessment	Student stumbles on some basic aspects of explanations and questions.
1	Disagree with student's assessment	Student fails to answer multiple basic questions. Lack of evidence that student performed work to acquire this skill.

The number of 3's, 2's and 1's a student receives from a grader will be counted up to give the numbers n_3, n_2 and n_1 respectively and then the percentages for this part of the assessment will be awarded as:

Percentage awarded	Criteria
20%	$n_3 \geq 2$
15%	$n_3 = 1$ **and** $n_2 = 2$
10%	$n_2 = 3$ or $n_3 = 1$ **and** $n_2 = 1$
5%	$n_3 = 0$ **and** $n_2 = 2$

2 Evaluation of quality of the project (40%)

In the grading we want to reflect that if you try a hard problem then innovation is not as vital while if you try an easy problem then you need to have a very good execution and/or you limit the maximum score you can get. An easy problem may correspond to a supervised classification problem on one easy dataset (i.e. train a ConvNet to solve MNIST and explore the property of depth). Note, however, you will be penalized if you tackle a very difficult problem (with lots of stages and moving parts) and then fail to execute. I have no problem if you initially consider a problem that turns out to be too difficult but then you pivot in time to tackle a more feasible aspect of the problem. So if you do not fulfill your initial goals that's fine, as long as you have good project management to consider what is feasible in the remaining time frame, the available computing resources and the group's skill levels.

You will be graded w.r.t. three categories

- Level of difficulty (s_D)

- Innovation (s_I)

- Execution (s_E)

For each category you will receive a score of between 1 and 3.

Score	Level of difficulty	Innovation	Execution
3	High	High	Good
2	Medium	Medium	Fine
1	Low	Low	Poor

These scores are then translated to a percentage as follows:

Percentage awarded	Criteria
40%	$s_D = 3, s_I \geq 1, s_E \geq 2$ or $s_D = 2, s_I \geq 2, s_E = 3$ or $s_D = 1, s_I = 3, s_E = 3$
35%	$s_D = 2, s_I \geq 2, s_E = 2$ or $s_D = 2, s_I = 1, s_E = 3$ or $s_D = 1, s_I = 2, s_E = 3$
30%	$s_D = 1, s_I \geq 2, s_E = 2$ or $s_D = 2, s_I = 1, s_E = 2$ or $s_D = 1, s_I = 1, s_E = 3$
20%	$s_D = 1, s_I = 1, s_E = 2$
10%	$s_D \geq 1, s_I \geq 1, s_E = 1$

2.1   Note on the level of difficulty score

The level of difficulty accounts for varying levels of ambition in the proposed projects. How will we assess the level-of-difficulty of a project? Undoubtedly, this will involve some level of subjectivity but we will give some guidelines in the *Project Ideas* document and in this subsection. We also hope in the feedback to your project proposal we will give an indication of how we view the level of difficulty of your proposed project.

To us the level-of-difficulty of the project depends on the coding effort required, the new technical/theoretical details that have to be understood and the difficulty in training and running the experiments. The assessment of the level-of-difficulty score will then be a trade-off between these three factors. So for example a coding effort involving an implementation from scratch would not require implementing an extremely difficult and complicated method to get a 3. Figure 1 gives an indication of the trade-off between the main three factors we will consider.
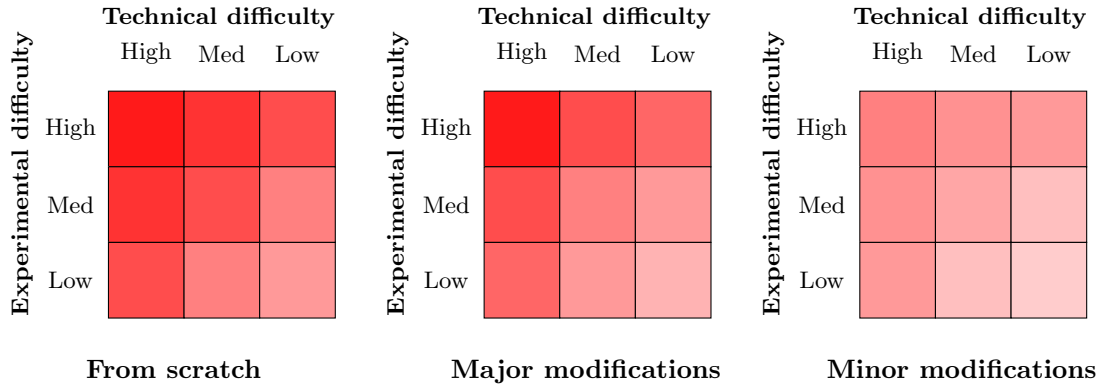
Figure 1: **Indication of how we will assign the level-of-difficulty score to your project**. We will consider three factors - the effort put into coding your implementation, difficultly of the theory behind the project and the effort needed to complete training and experiments. The colour in each cell of the matrix indicates the score you can expect to achieve. The redder the colour the higher the score. Each matrix corresponds to three crude categories of coding effort. The left matrix corresponds to the category when the group *implements everything from scratch* via a standard deep learning package, the middle matrix to when the group *downloads code and then adapts it somewhat significantly to facilitate an innovation and/or extensive ablation studies* and finally the right matrix to the case when the group *downloads code and then only makes very minimal changes*. Please note these matrices are guidelines as opposed an exact scoring scheme.

2.2   Note on the innovation score

The assessment of this part of the project is determined by

- how far you deviate from an implementation in a paper and/or

- are you applying a standard approach to a novel problem.

For the former you will not have to deviate too far from a paper to get a 3.

2.3   Note on the execution score

The assessment of this part of the project will come down to the set of experiments you run and the results you achieve. We do not expect everything you try to work but we do expect that you have put thought into the experiments you run and that you have done some testing to ensure you have bug-free implementations. Therefore execution refers, beyond the results obtained, to the design of the experiments to highlight the issues you want to investigate and also that you have given appropriate consideration to how long things will take to train given your resources.

3 Evaluation of the quality of the written report (20%)

Your written report will also count towards your final grade. Your report will be graded on these criteria:

1. Organization

2. Problem Statement / Motivation

3. Related Work

4. Presentation of technical material

5. Presentation and analysis of results

6. Discussions and conclusions

Once again you will be awarded a score of 1, 2 or 3 for each one of the 6 criterion. These scores will be added together to get s_r and then you will be awarded a percentage as follows:

Percentage awarded	Criterion
20%	$s_r \in [16, 18]$
15%	$s_r \in [13, 15]$
10%	$s_r \in [10, 12]$
5%	$s_r \in [6, 9]$

The scoring rubric for these criteria is available in appendix A.

4 Evaluation of the quality of the oral presentation (20%)

Your presentation will also count towards your final grade. Your group's presentation will be graded on these criteria:

1. Content and creativity

2. Coherence and organization

3. Speaking Skills and participation

Your group will be awarded a score of 1, 2 or 3 for criterion. The scores will be added together to get s_p and then you will be awarded a percentage as follows:

The scoring rubrics for these criteria are available in appendix B.

Percentage awarded	Criterion
20%	$s_p \in [8, 9]$
15%	$s_p \in [6, 7]$
10%	$s_p \in [4, 5]$

5   Final Grade

The percentage you achieve for each section *Skills acquired*, *Quality of project*, *Written report* and *Presentation* and the percentages from the assignment bonus points will be added together to get a final score s_f (2 bonus pts = 1%, capped at 10 bonus points). The final letter grade for your project will be computed based on this table

Grade	Criterion
A	$s_f \in [88, 100]$
B	$s_f \in [76, 87]$
C	$s_f \in [64, 75]$
D	$s_f \in [52, 63]$
E	$s_f \in [40, 51]$

Your final grade for the course will correspond to your project grade.

A Grading rubric of the written report criterion

Score	Criterion	Awarded when
3	Organization	Good organization; sections and arguments within sections are logically ordered; sharp sense of beginning and end.
2	Organization	Some organization; points jump around; beginning and ending are unclear
1	Organization	Poorly organized; no logical progression; beginning and ending are vague.
3	Problem Stat./Mot.	The projects objectives are clearly stated. The motivation for pursuing the project and its relevance are clearly and persuasively established by relating the project to related work. Supporting details specific to subject.
2	Problem Stat./Mot.	Some lack of clarity in objectives/purpose. The motivation for pursuing the project are somewhat clear, but no support provided.
1	Problem Stat./Mot.	Objective and motivation of the project are not clear or are not described.
3	Related Work	Prior work is acknowledged with appropriate references to theories, assumptions and prior findings.
2	Related Work	Some related work described, but unclear as to how they relate to the project or the link to the project is debatable.
1	Related Work	Little attempt is made to acknowledge the work of others. Most included references are inaccurate or unclear.

Score	Criterion	Awarded when
3	Pre. of tech. mat.	Clear presentation of the technical material central to the project and demonstrates it has been understood by the group members.
2	Pre. of tech. mat.	Somewhat clear presentation of the technical material central to the project and indicates it has been only partially understood by the group members
1	Pre. of tech. mat.	No or little technical detail presented and those that is presented has inconsistent or unclear notation.
3	Pre. & anal. of res	Clear description of the experiments run, the results and analysis of results, presenting not just raw evaluation data but conclusions that can be drawn from them. Relevant links are made between problem statement, claims made, and results.
2	Pre. & anal. of res	Somewhat unclear description of experiments run. Results are analyzed superficially with only raw numbers being dictated rather than potential causes for those results.
1	Pre. & anal. of res	No or little results shown. Very little analysis of results, with even a superficial analysis missing.
3	Dis. & con.	Discussion and conclusions attempt to tie the problem statement, experiments, and results together to tell an overall story.
2	Dis. & con.	Major components missing in the discussion. Little attempt to tie together experiments and problem statement/claims.
1	Dis. & con.	Little discussion or conclusions drawn.

B   Grading rubric for the presentation

B.1   Content and Creativity

Score	Meaning	Awarded when
3	Sufficient	The presentation contained relevant material clearly related to the main arguments. External research was used to justify arguments or solutions. The presentation of the material was presented in a creative way that held audience attention.
2	Lacking in some aspects	The presentation contained material to support the main arguments, but: 1) not all material clearly related to the main arguments; 2) limited external research was used to justify arguments or solutions; and/or 3) the presentation of the material was appropriate, but only somewhat held audience attention.
1	Insufficient	The audience had to make considerable effort to understand the underlying logical and flow of ideas. Major aspects of the analysis or recommendations were absent. No external research was used to justify arguments or solutions. The presentation lacked creativity and did not hold audience attention.

B.2 Coherence and Organization

Score	Meaning	Awarded when
3	Sufficient	The thesis, argument and solution were clearly stated and examples were appropriate. The transitions and flow were easy to follow. Slides were mainly error-free and logically presented.
2	Lacking in some aspects	The thesis, argument and solution were clearly stated, but: 1) not all examples were supportive illustrations; 2) the transitions and/or flow were some what difficult to follow; and/or 3) slides contained some errors but still presented logically.
1	Insufficient	The thesis, argument, solution and examples were not clearly stated. The conclusion was unclear. The transitions and flow were not logical. Slides contained errors and a lack of logical progression.

B.3 Speaking Skills and Participation

Score	Meaning	Awarded when
3	Sufficient	Group members were poised and had clear articulation. Every group member spoke and participated at a very high and balanced level. The presentation fit into the allotted time limit.
2	Lacking in some aspects	Team members were mostly audible and/or fluent on the topic, but:1) not all group members spoke and/or participated in a high and balanced level; and/or 2) the presentation slightly went over the allotted time limit.
1	Insufficient	Team members were often inaudible and/or hesitant. The presentation went well over the 10 minute allotment.