Assignment 4 - Project

Advanced Topics in Neural Networks

Project - 40 points

Choose one of the following projects:

1. Noisy CIFAR-100:
   https://www.kaggle.com/t/e95ff03613614b8f952fa8749ed22cd2
2. SemEval 2026 Task 5: Rating Plausibility of Word Senses in Ambiguous Sentences through Narrative Understanding:
   https://nlu-lab.github.io/semeval.html
3. SemEval2026 Task 3 DimABSA:
   https://github.com/DimABSA/DimABSA2026

Create teams of 1 to 3 students and participate in the chosen competition. Elaborate a research report in which you present the problem and related work, the solution you elaborated and your results.

Evaluation

You will be evaluated based on the following components. **All 3 components must be present to be graded**

5 points: You have to present your progress at least once in Week 10, Week 11 and Week 12.

35 points: The research report. The evaluation takes into consideration:

1. **Quality of the report**
2. **The Ablation Study**
3. **The Benchmark Performance**
4. **Oral presentation**: 10-15 minutes in Week 13 or Week 14.

During the presentation, you have to answer the following questions:

- Which aspect(s) influenced the best result, and why?
- Mention how could you improve the results, provided you had more time.

Quality of the report

- **Low quality:** for a simple description of related work, the new solution, and the results.
- **Medium quality:** for detailed related work, well-written solution description, presentation of results, and an analysis of them.
- **High quality:** an academic publication level of writing and structure in all aspects: related work, method description, theoretical analysis, results, and results analysis.

The Ablation Study

- **No ablation study:** no comparison between a baseline and the proposed solution.
- **Simple ablation study:** for basic comparison between the baseline and new solution.
- **Good ablation study:** for more detailed comparisons on various metrics such as speed, compute, memory usage.
- **Complex ablation study:** for extended comparisons in more scenarios, types of data, data regimes, evaluations of robustness, interpretability, and other more complex studies.

Benchmark performance

This section is graded based on the leaderboard of each competition. For the Noisy CIFAR-100 competition, the results will also be compared with previous years.