Assignment 14
LLMs
EECE-6822 Machine Learning
Cris Ababei
Electrical and Computer Engr., Marquette University

1. Objective
The objective of this activity is to run several code examples that study and illustrate various concepts related to LLMs.

2. Prerequisite Readings
Geron
- Ch.16: Natural Language Processing with RNNs and Attention

Raschka
- Ch.16: Transformers – Improving Natural Language Processing with Attention Mechanisms

3. Code Examples

Example 1: Attention mechanisms

This is the example code from **Ch.16** from Aurelian Geron's book.
[*B3-Geron] Aurelien Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly, 2022.
Open in your google colab and go through the code. You should read first the chapter itself from the book, before going through the code.
https://github.com/ageron/handson-ml3/blob/main/16_nlp_with_rnns_and_attention.ipynb
In this chapter, the author discusses how RNNs can be used to build an encoder–decoder architecture capable of performing neural machine translation (NMT), translating English to Spanish. Also, it explores attention mechanisms.

Example 2: Transformers ←

This code example is from **Ch.16** from Sebastian Raschka's book:
[*B3-Raschka] Sebastian Raschka, Yuxi Liu, and Vahid Mirjalili, Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python, Packt Publishing, 2022.
The source code (Jupyter Notebook) is located at the GitHub repository. You should read first the chapter itself from the book, before going through the code.
https://github.com/rasbt/machine-learning-book/tree/main/ch16
- This chapter explores the following topics:
- Improving RNNs with an attention mechanism
- Introducing the stand-alone self-attention mechanism
- Understanding the original transformer architecture
- Comparing transformer-based large-scale language models
- Fine-tuning BERT for sentiment classification

Example 3: Other tutorials

--What are Large Language Models
https://machinelearningmastery.com/what-are-large-language-models/
--A Gentle Introduction to Attention and Transformer Models

https://machinelearningmastery.com/a-gentle-introduction-to-attention-and-transformer-models/
--Datasets for Training a Language Model
https://machinelearningmastery.com/datasets-for-training-a-language-model/
--Positional Encodings in Transformer Models
https://machinelearningmastery.com/positional-encodings-in-transformer-models/
--Encoders and Decoders in Transformer Models
https://machinelearningmastery.com/encoders-and-decoders-in-transformer-models/
--Building a Transformer Model for Language Translation ←
https://machinelearningmastery.com/building-a-transformer-model-for-language-translation/
--Building Transformer Models from Scratch with PyTorch (10-day Mini-Course) ←
https://machinelearningmastery.com/building-transformer-models-from-scratch-with-pytorch-10-day-mini-course/

--10 Common Misconceptions About Large Language Models
https://machinelearningmastery.com/10-common-misconceptions-about-large-language-models/
--7 Concepts Behind Large Language Models Explained in 7 Minutes
https://machinelearningmastery.com/7-concepts-behind-large-language-models-explained-in-7-minutes/
--7 Advanced Feature Engineering Tricks for Text Data Using LLM Embeddings
https://machinelearningmastery.com/7-advanced-feature-engineering-tricks-for-text-data-using-llm-embeddings/

4. **Assignment**

Modify the code example from this tutorial:
"Building a Transformer Model for Language Translation"
https://machinelearningmastery.com/building-a-transformer-model-for-language-translation/
(which currently translate English - French) to develop a different translation model.
Select your own choice of dataset to work with from:
https://www.manythings.org/anki/
Investigate three different Transformer architectures.

5. **Deliverables**

You must write (typed) a report and upload it as a PDF file on D2L. The report should be named "**hw14_report_LastName.pdf**". You should also create a .zip archive with all your code and implementations of all parts of the assignment. Upload also this archive .zip file with the name "**hw14_implementation_code_LastName.zip**" to D2L. Hence, your D2L should contain two items: the report and the .zip file. **Do not include the report inside the .zip and upload only the .zip. They should be two separate items!**

The report should include the following sections and subsections. Make sure section titles are in bold font and pages are numbered.
1) **Title + course info + your name**
2) **Summary.** Describe in one paragraph what the objective of the assignment is.
3) **Description of Experiments and Discussion.** Describe the experiments you did. All tables and figures should be numbered and should have captions. All plots in all figures should have axes labels and titles. Present a meaningful discussion with the interpretation of the results you obtained. Explain if you expected the results or not; discuss the intuition behind it.

4) **Conclusion.** Present your conclusions; highlight what are your main takeaways that you learned from this assignment. Describe what issues you encountered and how you solved them.
5) **References.** Include all references that you used, as a numbered list. Cite them in the report itself; do not just list them! If your report has References that are not numbered and cited in the report, points will be deducted!