Assignment 2

(Due: Nov 9, 06:00am)

Zoie loves movies! *Ready Player One* by Steven Spielberg, for example. She often checks movie review websites (like many of us) to decide if she would like to watch the movie in the theatre or rather wait a few months and watch it online.

One day she visited HKU again and came across a student who happens to study COMP3361 this semester, again.

"We learned a NLP application called *sentiment analysis* recently!", said the student with excitement.

"That's nice! What else you've learned?", asked Zoie, with a warm smile on her face.

"Parsing, but it's rather boring", answered the student.

"Okay." Zoie started to giggle, "Maybe you can make it useful somehow. For example, can you build a sentiment analyser using parse trees to help me to choose movies based on people's reviews?"

"Sure! Anything!", said the student.

And here comes the **assignment 2**.

This assignment is based on the Stanford sentiment treebank https://nlp.stanford.edu/sentiment/index.html and the pytorch example code https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html.

Your task is to adapt the code in the tutorial into an sentiment analyser. Don't worry about the performance, we won't evaluate on that. It's just a toy neural network and a toy example. In fact, do not change the size of the model or the number of training iterations. Those are just there to make sure the code is running. All tasks are labeled as TODO in the colab.

- Task 1: Write a function to extract the words from the tree string. (10%)
- Task 2: Adapt the LSTMTagger code into a LSTM for the sentiment analysis model. (50%) Task 2 has three sub-tasks with 10%, 20%, 20%, respectively.
- Task 3: Adapt the training loss. (15%)
- Task 4: Write the evaluation and report the accuracy of your model on the dev set. (20%)
- Task 5* (open task): Can you use the tree structure there to further improve the model? Write the code and explain. (5%)

You do not need to write a report, add text and comments in the colab to let us know what you did. Submit your colab as a .IPYNB to moodle. Late submissions will not be graded.

Set back and relax, Zoie. The movie is about to start!