SOI1010 Machine Learning II - Assignment #3

Due: December 8, 2023 11:59 pm

You know the drill. The submission should include a code (both link to the colab and .py format) and a report that has answers to the questions and results. Use ***PyTorch*** (or TensorFlow/JAX). Also, minimize the use of numpy. There will be mark deductions if numpy is used when PyTorch is preferable or should be used. Marks will be deducted if the submission does not include the requested files. **DO NOT** use other libraries, such as scikit-learn/sklearn, to use a model you are supposed to implement. **Using sklearn or any other third library or already built-in functions** that you are asked to implement **will result in 0 mark**. Also, if an assignment asks you to implement some model, that means you shouldn't use the built-in implementation from any library for that model in the first place.

NEW guidelines:
1. Do NOT copy codes from references/online resources, especially from DACON/Kaggle. If you get caught copying/combining parts of the codes that are being shared online, you will get 0.
2. Do not use test data (from test.csv) for training your model. Test data should only be used for test. If the results are not reproducible, you will get 0.
3. Do not use external data. Stick with what you are provided with from Kaggle competition.
4. Do not use a pre-trained model. You should train a model from scratch.
5. In other words, use only train data (from train.csv) for training your model.

Kaggle competition: Sign Language Recognition

a) The competition/dataset can be found here.
b) Your task is to achieve as high performance as possible (without cheating, of course). You can use any model you want. But, I suggest that you start with MLP and Convolutional Neural Networks (CNN). I strongly encourage you to research and study advanced neural network models and implement them to get better performance.

[Bonus/optional] Explore

a) If you have finished the above two with ease and crave for more challenges, you are welcome to explore more datasets/competitions from DACON and Kaggle. Take a look at the current and trending challenges. Then, you can get an idea as to which problems are deemed important to solve for our society.
b) Then, think about which problems you find interesting. Better yet, you can also come up with new problems. Think about the challenges that you may encounter for tackling the problems, such as data collection, modeling, etc.
c) If you are more interested in pursuing research problems, you can also contact me via e-mail.
d) You should continuously embark on this kind of journey of exploration, not just during the semester, but also during the breaks. Actually, throughout the undergraduate years. In the end, you are a prospective data scientist aspiring to be one of the best in the world (right?). So, go make a team to participate in competition/hackathon/etc. with challenges you want to tackle, gain experience, and improve your skills. Alternatively, join a research group led by a professor who pursues research fields of your interest if you are more of a researcher person. In process, you get to realize what you want to do, what you are passionate about, what you are good at, and most importantly, what kind of a person you are.