CSCI 4050U / 5070G
Machine Learning
Final Project, 2023

The final project is a collaborative team exercise.

- Teams with a maximum size of three.
- Submit:
    - PDF of a Jupyter Notebook / Google Colab
    - PDF of presentation slides
    - A demonstration YouTube video of at most 5 minutes. The video can be a slide show, a demonstration of the neural network, or a combination of both. Elements of the video should include the description of the learning problem, elements of the neural network architecture, and the deployment of the neural network solution.

**Project Description**

In this project, you are to work with a dataset of your choice, and formulate a well motivated problem that can be addressed by a machine learning solution.

**Dataset**

You can collect data using any means including but not limited to the following data sources:

1. Standard datasets provided by PyTorch:
    a. https://pytorch.org/vision/stable/datasets.html
    b. https://pytorch.org/text/stable/datasets.html
2. Kaggle datasets that are publically available.
3. Open data repositories available at various government agencies including Canadian Open Data portal, Statistics Canada, Ontario Open Data, or any others you find.

**Problem**

You must clearly describe your learning problem in the following contexts:

1. Describe the learning problem in the Jupyter Notebook.
2. Describe the learning problem at the beginning of the video.

The types of problems that are suitable for the project include but not limited to the following:

1. A comparative study of different machine learning approaches
2. An experimental study of a novel neural network architecture
3. The training and novel application of the resulting network

**Neural Network Design**

You are at the liberty of utilizing any neural network architecture. I recommend that you keep the architecture reasonably simple given the limited time and possible computational resources. Be aware of the potential computational costs (time and CPU time) associated with the training. Use the model size (number of parameters).

One standard technique to improve the expressive power of the network is to rely on pre-trained models as an initial network, and use limited computation to fine tune for your specific training data.

**Training and Deployment**

Going beyond the Jupyter/Colab worksheets, you are expected to investigate how the trained neural network can be integrated into applications and systems. Include in your presentation a demo of the deployment of the trained network in action. You are expected to demonstrate a running system that uses the trained neural network.

**Presentation**

You must prepare three forms of presentation:

1. PDF export of the Jupyter/Colab worksheet. This does not need to include deployment code. The worksheet only needs to include the

neural network architecture and basic training loops, and proof-of-concept validation and test results.

2. A presentation (Google Slides/PowerPoints/Key Notes) on the project.
3. A YouTube video (**max 5 minutes**) presenting and demonstrating the running of your system.

**Submission**

You are to submit your work via the following Google Form:
https://forms.gle/RVz5WASKrKuDf9167

You are required to enter:

1. Project title
2. All team members by their emails
3. Upload Jupyter or Colab worksheet (PDF export)
4. Upload presentation slides (PDF export)
5. Specify the presentation YouTube video URL

The grading rubric is:

| | |
|---|---|
| Dataset collection | 10% |
| Problem formulation | 20% |
| Neural network design | 20% |
| Training and deployment | 20% |
| Presentation slide quality | 15% |
| Video presentation quality | 15% |

**Sample Projects**

1. Learn the theory and implementation of a new neural network architecture that has not been covered in class.
    a. Autoencoder and variational autoencoders (VAE)
    b. Generative Adversarial Networks (GAN) and image generation

2. Perform data challenges on Kaggle using neural network based solutions.
3. Investigate different design choices and compare them on a common dataset.