Assignment 3: Deep Learning

Submission: Monday June 3rd
3 students per group

Prof. Fabio A. González
Machine Learning - 2024-I

Download the dataset MM-IMDB from `https://www.kaggle.com/datasets/johnarevalo/mmimdb`. The dataset includes poster images from movies of different genres as well as the text of the corresponding synopses.

We will train a metric learning model that learns a metric over movie posters that reflects whether two posters belong to the same or different movie genres. A movie may belong to more than one genre, e.g. it may be both an *animation* and *adventure* movie. To compare two movies x_i and x_j with labels y_i and y_j we can use the Jaccard index, $J(y_i, y_j)$. $J(y_i, y_j) = 1$ if x_i and x_j have the same labels (share the same movie genres) or $J(y_i, j_j) = 0$ if they do not have any common label. The goal is to be able to calculate this metric but only using the visual information in the poster. The approach that we will follow is metric learning using a triplet network.

1. **Build the triplets training/validation dataset**

   Triplet networks require triplet samples to be trained. A triplet sample comprises three elements an anchor (x), a positive sample (x^+) and a negative sample (x^-). The positive sample is expected to be most similar to the anchor than the negative sample. Triplets can be selected at random but this does not work very well. It is better to have a smarter sampling strategy.

   (a) To build the dataset follow the next procedure:

   - Repeat m times:
     i. Select at random a batch B of b of images from the training partition
     ii. Repeat n times:
       A. Select one image x_i from B at random
       B. Sort all the remaining $b-1$ images according to the label Jaccard index against y_i
       C. Choose the most similar movie as the positive sample, x_i^+, and the least similar movie as the negative sample, x_i^-.
       D. Store the triplet (x_i, x_i^+, x_i^-) along with the Jaccard distances $1 - J(y_i, y_i^+)$ and $1 - J(y_i, y_i^-)$

   (b) Show some of the triplets from the dataset displaying both the poster and the labels of anchors and samples.

2. **Train a triplet network**

   (a) Adapt this example using the log-ratio loss function described in this paper.

   (b) Train the network with the training/validation set built in the previous item.

   (c) Show the output of the model for some of the triplets in the validation dataset. Discuss.

(d) Select at random 1,000 movies from the test dataset. For each movie calculate the corresponding embedding using the learned model. For some of the movies calculate the 10 nearest neighbors using the Euclidean distance between the embeddings. Show the nearest neighbors posters along with their labels. Discuss the results.

(e) When comparing the performance of different models you may use the average discounted cumulative gain over a set of query validation images and the corresponding retrieval results.

3. **Embedding visualization**

(a) Use t-SNE to visualize the test dataset.

(b) Draw different plots to show the distribution of different genres. Discuss.

The assignment must be submitted as a Jupyter notebook through the following Dropbox file request, before midnight of the deadline date. The file must be named as `ml-assign3-unalusername1-unalusername2-unalusername3.ipynb`, where `unalusername` is the user name assigned by the university (include the usernames of all the members of the group). You do not need to submit additional files all the detail must be included in the notebook. Make sure that the notebook renders correctly and is free of errors before submitting.