**Advanced Machine Learning**

**Coding Assignment 1**

1. The Blog Authorship consists of the collected posts of 19,320 bloggers gathered from blogger.com in August 2004. The corpus incorporates a total of 681,288 posts and over 140 million words - or approximately 35 posts and 7250 words per person.

Each blog is presented as a separate file - and labeled for gender and age.

Here is a link to the corpus.

<https://u.cs.biu.ac.il/~koppel/BlogCorpus.htm>

**Task**

Implement a CNN based NN model to classify the gender and age of a blogger. Use standard k-fold cross validation. Use dropout for training.

**For testing – new data will be used by instructors. Keep provision for testing**

1. Open Images is a dataset of ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localized narratives. It contains a total of 16M bounding boxes for 600 object classes on 1.9M images, making it the *largest existing dataset with object location annotations*. The boxes have been largely manually drawn by professional annotators to ensure accuracy and consistency. The images are very diverse and often contain complex scenes with several objects (8.3 per image on average).

<https://storage.googleapis.com/openimages/web/factsfigures.html>

**Task**

Implement a CNN based NN model to perform the two tasks – (a). Image classification and (b). Object recognition. Use standard k-fold cross validation. Use dropout for training.

Note – you can use a single multi-objective architecture or implement two different independent architectures for the two tasks.

**Submission**

Should contain all details about the final architecture(s), training, accuracy drops and time for training observed with different dropout parameters during training and final accuracy results.

**For testing – new data will be used by instructors. Keep provision for testing**