COM6509 Assignment 2 - Deadline: 11:00 AM, Friday 22nd Jan 2021

Click for FAQ at the bottom of this document.

A. Assignment Brief

Please READ the whole assignment first, before starting to work on it.

A1. How and what to submit

1) A **Jupyter Notebook** with the code in **all** the **cells executed, output displayed,** and **code documented.**
2) Upload your notebook to Blackboard before the deadline above. Name your file as **COM6509_Assignment_2_USERNAME.ipynb**, where USERNAME should be replaced with your username such as **abc18de**
3) **NO DATA UPLOAD**: Please do not upload the data files used. We have a copy already. Instead, please use a **relative file path in your code (data files under folder 'data')**, as in the lab notebook so that we can run your code smoothly when needed. So '**./data/**', instead of '**/User/username/myfiles/mlai/assignment1/**'

A2. Assessment Criteria (Scope: Sessions 6-8; Total marks: 30)

1) Being able to build complete, reproducible machine learning pipelines from loading data to evaluating prediction performance.
2) Being able to design different machine learning models to compare/optimise prediction performance.
3) Being able to perform exploratory data analysis to gain insights.

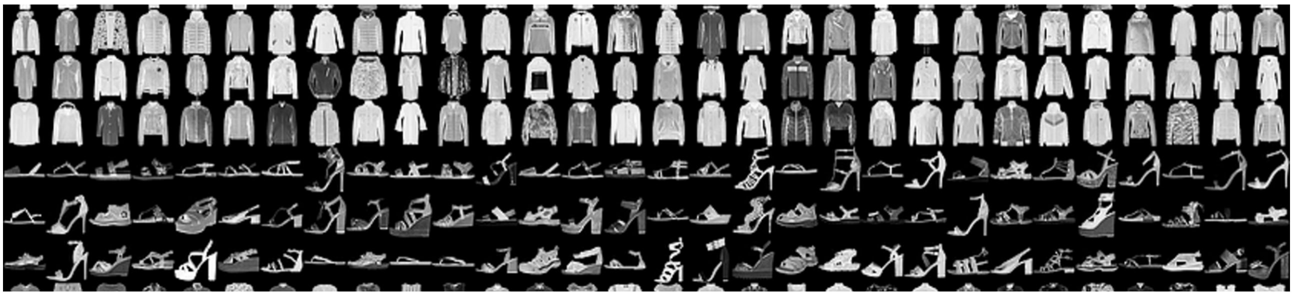
A3. Late submissions

We follow the Department's guidelines about late submissions, i.e., a deduction of 5% of the mark each working day the work is late after the deadline, but **NO late submission will be marked one week after the deadline**. Please see this link.

A4. Unfair mean

Please carefully review the handbook on what constitutes Unfair Means if not sure.

B. Assignment on Fashion-MNIST [30 marks]



Fashion-MNIST is a dataset of Zalando's article images, with examples shown above. It consists of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, associated with a label from 10 classes: *0=T-shirt/top; 1=Trouser; 2=Pullover; 3=Dress; 4=Coat; 5=Sandal; 6=Shirt; 7=Sneaker; 8=Bag; 9=Ankle boot*. In this assignment, we will work on this dataset.

- You will make several design choices (e.g. hyperparameters) in this assignment. There are no "standard answers". You are encouraged to explore several design choices to settle down with good/best ones, if time permits.
- The assignment questions specify the tasks and you make your design choices to complete the tasks. You are free to use scikit-learn or pytorch, batching or no batching, as long as you can complete the assignment questions.

B1. Reproducibility & readability [2 marks]

Whenever there is randomness in the computation, you MUST set a random seed for reproducibility. Use your UCard number XXXXXXXXX as the random seed throughout this assignment. [1 mark]

Answers for each question should be clearly indicated in your notebook, e.g., including question numbers below in bold such as **B2.1a**. All code should be clearly documented and explained. [1 mark]

B2. Supervised learning on Fashion-MNIST [16 marks]

We aim to train machine learning models to classify the 10 classes in Fashion-MNIST using the standard train/test split with decent performance, i.e. much better than the chance level at worst.

B2.1 Data loading and inspection [2 marks]

a) Use the PyTorch API for Fashion-MNIST to load both the training and test data of Fashion-MNIST. You may refer to similar procedures in Lab 7 for CIFAR-10. Preprocessing is NOT required but you are encouraged to explore and use preprocessing such as those in the torchvision.transforms API. [1 mark]
b) Display at least eight images for each of the 10 classes (8x10=80 images). [1 mark]

B2.2 Evaluation metrics [2 marks]

Keep a record of the four metrics M1-4 below for each of the six models in B2.3 and B2.4:

M1) Training accuracy: the prediction accuracy of the trained model on the training dataset.

M2) Testing accuracy: the prediction accuracy of the trained model on the test dataset.

M3) Training time: the time taken to train the model (i.e. to learn/estimate the learnable parameters) on the training dataset.

M4) The number of learnable parameters of the model.

B2.3 Logistic regression [4 marks]

If a hyperparameter needs to be set, you are free to choose one that can deliver satisfactory results for you.

a) Train a logistic regression model on the training set of Fashion-MNIST and test the trained model on the test set of Fashion-MNIST. Report the four metrics M1 to M 4 and plot a confusion matrix for predictions on the test data. [2 marks]
b) Train and test a logistic regression model with L1 regularisation as in a). Report M1 to M4 and plot a confusion matrix for predictions on the test data [1 mark]
c) Train and test a logistic regression model with L2 regularisation as in a). Report M1 to M4 and plot a confusion matrix for predictions on the test data [1 mark]

B2.4 Convolutional Neural networks (6 marks)

This question asks you to design various convolutional neural networks (CNNs). Only the number of convolutional (Conv) layers and the number of fully connected (FC) layers will be specified below. The CNN in Lab 7 can be a reference but you are free to design other aspects of the network. For example, you can use other types of operation (e.g. padding), layers (e.g. pooling, or preprocessing (e.g. augmentation), and you can choose the number of units/neurons in each layer. Likewise, you may choose the number of epochs and many other settings according to your accessible computational power. Reminder: there are no standard answers.

a) Design a CNN with two Conv layers and two FC layers. Train and test it as in B2.3a. Report M1 to M4 and plot a confusion matrix for predictions on the test data. [2 marks]
b) Design a CNN with two Conv layers and five FC layers. Train and test it as in B2.3a. Report M1 to M4 and plot a confusion matrix for predictions on the test data. [2 marks]
c) Design a CNN with five Conv layers and two FC layers. Train and test it as in B2.3a. Report M1 to M4 and plot a confusion matrix for predictions on the test data. [2 marks]

B2.5 Performance comparison (2 marks)

a) Summarise each of the four metrics from the six models in B2.3 and B2.4 using a bar graph. In total, four bar graphs need to be generated and displayed, one for each metric with six results from B2.3 and B2.4. [1 mark]
b) Describe at least two observations interesting to you. [1 mark]

B3. Unsupervised learning on Fashion-MNIST [12 marks]

Choose two out of the 10 classes according to your preference and use **only the training data for these two chosen classes** to complete all tasks in this section B3. It will be better to finish reading the remaining part of this section before choosing the two classes. Again, you may choose any two and there is no “correct” answer about which two to choose but some choices may make your studies below more interesting than others.

B3.1 PCA and *k*-means [7 marks]

a) Apply PCA to all images of these two chosen classes. Visualise the top 24 eigenvectors as images and display them in the order of descending corresponding values (the one corresponding to the largest eigenvalue first). [2 marks]
b) Use the top 24 PCs to reconstruct 30 images, with 15 from each class (any 15 images are fine from each class). Compute and report the mean squared error between the reconstructed and original images for these 30 images (a single value to be reported). Show these 30 pairs of reconstructed and original images. [2 marks]
c) Plot the PCA representations of all data points in a 2D plane using the top two PCs. Use different colours/markers for the two classes for better visualisation (*Hint: You need to use the class labels here for visualisation*). [2 marks]
d) Use *k*-means to cluster all data points as represented by the top two PCs (clustering of two-dimensional vectors, where each vector has two values, PC1 and PC2). Visualise the two clusters with different colours/markers and indicate the cluster centers clearly with a marker in a figure similar to question c) above. [1 mark].

B3.2 AutoEncoder [4 marks]

a) Design a new autoencoder with five Conv2d layers and five ConvTranspose2d layers. You are free to choose the activation functions and settings such as stride and padding. Train this new autoencoder on all images of these two chosen classes for at least 20 epochs. Plot the mean squared error against the epoch. [2 marks]
b) Modify the autoencoder in 3.2a so that the code (bottleneck) has a dimension of 2 only. Plot the 2-dimensional representations in terms of this autoencoder code for all data points in a 2D plane as in 3.1c and cluster them as in 3.1d, showing similar colour/marker visualisation. [2 marks]

B3.3 Observation [1 marks]

Describe at least two observations interesting to you from B3.1 and B3.2 above.

The END of Assignment

C. FAQ or further clarification

1. The questions mean what they ask. When B2.2. M1 says to report accuracy on the training set. It means the training set. When B3 says to use only the training data. It means to use only the training set for B3.
2. Lab 7 CNN has three FC layers fc1, fc2 and fc3. Another example explicitly states this is slide 35 of lecture6 about AlexNet, which has three fc layers too.
3. You can use Seaborn for plotting the confusion matrix. See an example in my previous year's Lab 8 notebook (on Naive Bayes).
4. Data conversion between pytorch and numpy: see "Converting NumPy Array to Torch Tensor" and "Converting a Torch Tensor to a NumPy Array" in pytorch tensor tutorial
5.