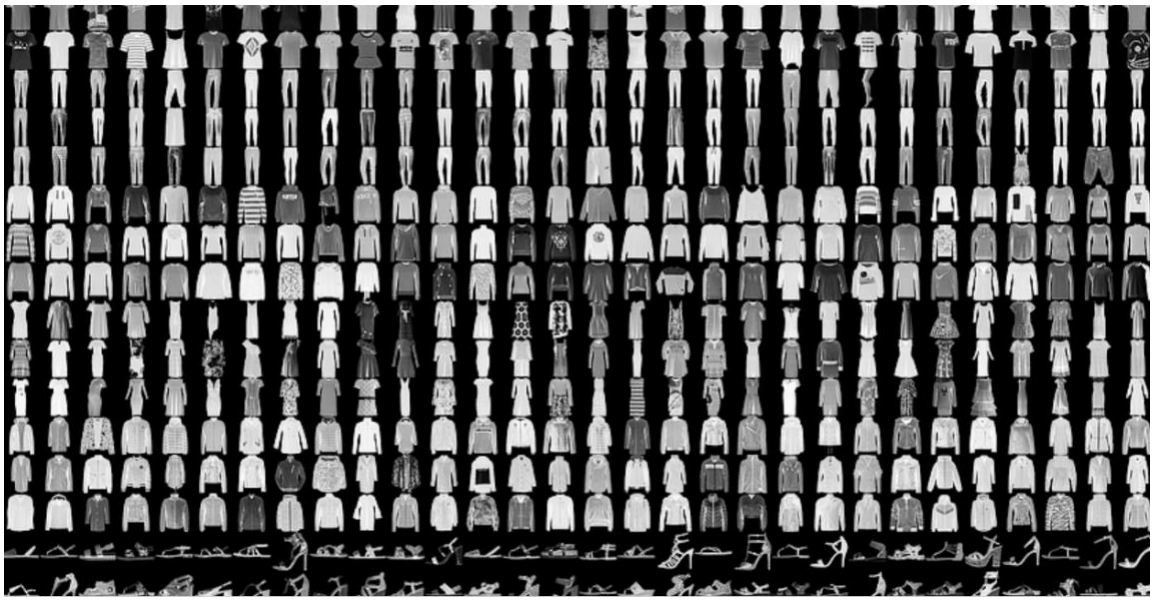
BIRZEIT UNIVERSITY

Electrical and Computer Engineering Department

ENCS5341 – Machine Learning and Data Science

Project – Fashion-MNIST

Submission Deadline: Feb 10, 2023
Team Work: Teams of two students.



Sample images from the Fashion-MNIST dataset.

In this project, we will test machine learning models for classifying images of clothes from the Fashion-MNIST dataset. The Fashion-MNIST dataset 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.

You need to complete the following tasks:

1. Validation Set

Create a validation set by splitting the 60,000 training examples into 40,000 examples for the training set and 20,000 examples for the validation set. The splitting of the examples must be random and make sure to shuffle the examples before creating the splits.

2. Features extraction

Features extraction requires domain knowledge. To make things simple, we will use the raw pixels of the images as features-vector. In other words, each image will be represented by a 784-dimensional features vector, where the first 28 elements in the vector are the 28 pixels of the first row in the image, and the elements in column 29 to column 56 are the 28 pixels of the second raw in the image, and so on. Keep in mind that you might need to normalize your features.

3. Baseline Model

As a baseline model, evaluate a nearest neighbor baseline. For the baseline, you have to decide between two distances of your choice and you need to set k to either 1 or 3. In total, you will have 4 different combinations of the distance measure and k. Choose the best combination by evaluating all your choices on the validation set. Report the accuracy of the chosen baseline on the test images.

4. Improvements

Try to achieve better performance by evaluating two additional models on the task. Discuss and motivate your model selection, and comment on why the performance has improved (or potentially did not improve). If the chosen models have hyper-parameters, make sure to tune at least one hyper-parameter for each model using the validation set. To tune a parameter, you need to test at least three different values.
For each model report the accuracy on the training, validation, and testing sets.

5. Error and Evaluation Metrics Analysis

Compute the accuracy of the best model from the previous part on each class separately using the test images. For the class with the lowest accuracy, select 20 random images where the prediction is wrong and analyze the predictions of your model. Do you notice any interesting pattern? If yes, can you use your analysis results to improve your model?
Finally, discuss whether the accuracy is a good metric for the task. If not, what would be a better metric?

6. Different Features

Try to use another (probably better) features of your choice (do your research) and re-evaluate both the baseline and your best model using the new features. Discuss the reasoning behind your features selection.

7. Report

Besides your code, you need to submit a report (max. 8 pages) to analyze and discuss your results. Your project will be mainly evaluated based on the quality of the report (60% of the project grade). The report should have the following sections:

- Introduction: describe the task you are addressing in the project and the different models you have tried. Also, define the evaluation metric that you have used.
- Technical Details: for each model you have tried, list the values of all the hyper-parameters you have used and explain the meaning of each one.
- Experiments and Results: discuss all the experiments you have done (model evaluation, hyper-parameters selection, …. etc) and the reported results in each case. Comment on all the results.
- Conclusions and Discussions: discuss your conclusions and also comment on the limitations of the used models and the evaluation metric.