MATH498: Mathematical Foundations of Machine Learning
Project 2: Multiclass classification

Submission info:

- Upload report on gradescope
- Send me the code or link your (public) github on the report
- Upload predictions on Kaggle

Introduction

In this project you will explore the capabilities of neural networks.
The dataset considered is the Fashion NMIST `https://github.com/zalandoresearch/fashion-mnist`.
Similar to the NMIST dataset we saw in the class, which is often used for multiclass classification task, where a handwritten number is labelled with the number it represents, this dataset contains images of clothing items, and their labels.

Dataset description

The dataset contains in total 70,000 square images of 28x28 pixels and respective labels.
Each image contains a picture of a clothing item (e.g. trousers) and its corresponding label (trousers).
There are 10 possible classes.

Submission

- Submit your predictions on the validation set and holdout set
- A brief report that explains how you obtained your results. A template for LATEX in the file "report.tex" is provided. (If you do not want to use LATEX, please use the same sections as shown in "report.pdf".)
- Upload code files with the report (as a PDF file) along with your code/notebook or parameters/screenshots of the tools you used. For further instructions refer to the report template.

We might ask you to show us what you did, so please keep the necessary files until the end of the semester.

Grading

The report counts for 20% of the grade. It will be judged on the following criteria:

- Does the experimental set-up make sense? (e.g.: does the report have all the necessary sections, are the performance metrics listed)
- Do the conclusions/lessons learned make sense?
- Was cross-validation was performed for model selection / hyper parameter selection?

- Are engineering choices (e.g. feature selection, hyper-parameter selection, model selection) justified?
- Does the report contain enough information to reproduce the experiment?

The rest of the grade is computed the following way.
Each submission (upload of a prediction file for a given data set) will be ranked according to the accuracy of the predictions. The average accuracy is defined as:

$$\text{accuracy} = \frac{1}{k}\sum_{k=1}^{K}\frac{tp_k + tn_k}{tp_k + fp_k + tn_k + fn_k},$$

where K denotes the number of classes, tp_k the number of true positives, tn_k the number of true negatives, fp_k the number of false positives, fn_k the number of false negatives, all with respect to class k.
Now we compare the cost of the submission to two baseline predictions: a weak one (called "baseline easy") and a strong one (called "baseline hard"). These will have a cost of CBE and CBH respectively, calculated as described above. Both baselines will appear in the rankings together with the error measure of your submitted predictions. Performing better than the weak baseline on the public set will give you 50% of the grade, and matching or exceeding the strong baseline on the hidden set will give you 80% of the grade. This allows you to check if you are getting at least 40% of the grade by looking at the ranking. If your prediction performance on the hidden set is in between the baselines, the grade is computed as:

$$Grade = \left(1 - \left(\frac{Cost_{test} - CBH_{test}}{CBE_{test} - CBH_{test}}\right)\right) \times 40\% + 40\%$$

$CBE_{test} = 0.7$
$CBE_{hard} = 0.9$