**Image Classification of the Fashion MNIST dataset**

**A machine learning project report**

**BSMALEA1KU**

|  |  |
| --- | --- |
| Gabriel Cristian Circiu Maciej Pawel Jalocha  Wenzel Keil | gaci@itu.dk  macja@itu.dk  weke@itu.dk |

# **Introduction**

This section introduces the topic and provides background information.

* Introduce the dataset and what it contains, where it’s from, release time etc.
* How was it split up? Mention 10k train and 5k test data.
* Any observations on it? Perhaps a plot of distribution of classes.
* Was there any related work done already to study?
* Any assumptions we might have and initial insights.
* Any data manipulation we did, like for contrast etc.

Required: Exploratory data analysis. The report should introduce the reader to the data by illustrating selected aspects of the data.

Required: Visualisation of data. As part of the exploratory data analysis our report should present a visualisation of the dataset based on dimensionality reduction obtained from PCA. Remember to include some comments on what the reader can learn from our visualisation

# **Methodology**

This section describes the methods and techniques used in the project.

* Write about all 6 methods, DT (from scratch/library), FFNN (from scratch/library), RF, Template Matching
* Did we implement them the way it should be? Using learned methods from lectures.
* Cite the libraries used for the library implementations
* Perhaps show how a calculation looks like in code, or at least the formulas themselves
* Source on where all the code can be found online (our github repo)
* Hyperparameters we use or changed, i.e. special gradient descent in FFNN
* Explain that our custom template matching has no previous implementation anywhere so we cant evaulate correctness on that
* Mention cross validation to ensure that we are not seeing an outlier result

Required: Details on implementations. Your report should describe and discuss the key points of how you implemented the two methods, M1 and M2. Please also include a discussion of how you have asserted your implementation’s correctness.

Required: Details on machine learning methods. For each method, please make sure to include a description of how you applied the method to the data including details needed for an independent reproduction of your results and a discussion on how you have gone about selecting any hyperparameters for the method.

# **Results**

This section presents the results of the project.

* Discuss results of all methods, and all the results, most importantly accuracy and auc.
* Loss/Cost function results over epochs
* Compare what do the results show and what they mean, which one we foun to be most important
* Random forest good, but neurla network even better
* Explain the difference our feature extractions made
* Explain the difference of using PCA

Required: Interpretation and discussion of the results Your report should include a thorough discussion of the performance of each of the methods applied. In particular, you should compare the methods’ performance and guide the reader in interpreting the results. Use your expert knowledge to explain the results; for instance, why do particular methods perform better than others?

# **Discussions**

This section discusses the implications of the results.

* What do the results mean?
* Are our models good?
* Are there any metrics that show areas that could be improved?
* Would these models be qualified to be used in the real world?
* PCA helps reduce computation required
* Extracting high level features could also help

# **Conclusion**

This section concludes the report and suggests future work.

* Using different neural networks
* Neural network is the right approach
* Scale it up to other datasets