# Foundations and Frontiers of Machine Learning Group Assignment 2 Group 1

Contents

[Foundations and Frontiers of Machine Learning Group Assignment 2 Group 1 1](#_Toc179642033)

[1. Table of Contributions 1](#_Toc179642034)

[2. Data visualization (Task 1) 1](#_Toc179642035)

[3. Perceptrons (Task 2) 2](#_Toc179642036)

[4. Multilayer Perceptrons (Task 3) 2](#_Toc179642037)

[5. CNN (Task 4) 2](#_Toc179642038)

[6. Visualising CNN Outcomes (Task 5) 2](#_Toc179642039)

[7. Multitask learning – Fashion MNIST (Task 6) 2](#_Toc179642040)

## Table of Contributions

1. Data visualization - Emmanouil
2. Perceptrons – Emmanouil
3. Multilayer Perceptrons – Emmanouil
4. CNN
5. Vicualising CNN Outcomes

## Data visualization (Task 1)

Machine learning models use data to finetune the value of their internal parameters in a process called training. In order to produce a model with high classification or regression ability, the dataset must contain information related to the specific task, thus bigger datasets usually result in better performance of the same models. Not all information within the dataset is useful, however. Oftentimes, a feature contains information that has already been made available by another feature or a combination of features, and as number of features grows larger (as is the case usually for bigger datasets), the chances that a feature consists of the linear combination of other features of the dataset increase. At the same time, books or electronic devices humans read from are best suited for up to two features, as they physically have two dimensions. Using techniques such as video or colour/shape encoding can increase that by some amount, but datasets can have dimensions in the order of millions, orders of magnitude higher than what humans can visualise or comprehend.

A diagram of a colorful circle

Description automatically generated with medium confidenceThe issue of repeated information within the dataset can be addressed by finding new features that are independent from each other but contain the same information, with the Principal Components Analysis (PCA) technique. The first new feature is engineered to contain as much variance as possible from all features of the dataset. Each next one is engineered to contain as much of the variance that is left, without incorporating any information of the previously engineered features. This often means that the last features contain little to no additional information (depending on how linearly independent the dataset was) and can be discarded, allowing for easier visualization and reducing the overall noise. What is more, the features are now in descending order in terms of information included so we can select the first n ones, knowing that we have maximum information per feature density. We can now select a number that is suitable for visualization. The new features have no actual meaning but can be used to show the distribution of the samples in the space, informing us about the separability of the classes. If we plot the first two components with different colour for each class, we can see which classes overlap, and which ones are linearly separable. Thus we have that, for example, the pairs 0 and 1, 0 and 2, 2 and 9 are some of the most easily to differentiate since the occupy opposing spots on the plane. Plotting the classes in pairs would naturally offer a more clear perspective as some of the classes are now hidden behind others.

*Fig 1. PCA plot of MNIST handwritten digits dataset for first 2 dimensions.*

## Perceptrons (Task 2)

asdfasdf

## Multilayer Perceptrons (Task 3)

asdfasdfasdf

## CNN (Task 4)

asdfasdfasdfasdfasd

## Visualising CNN Outcomes (Task 5)

asdfasdfasdfasdfasdfasdfasdf

## Multitask learning – Fashion MNIST (Task 6)

asdfasdfasdfasdfasdfqwerqwerqwer