MULTILAYER PERCEPTRON

DMML ASSIGNMENT-3

Date:	21-04-2024
Authors:	Hiba AP -MDS202326
	Nooh Ali –MDS202337

INTRODUCTION

Here we are trying to replicate a semi-supervised learning experiment demonstrated using K-Means clustering on the MNIST dataset. We extend this to Fashion MNIST and Overhead MNIST datasets. Starting with 50 clusters, we experiment with varying K values to examine semi-supervised learning's efficacy. This brief aims to explore how different cluster sizes influence MLP model performance in semi-supervised learning for image classification tasks.

DATA SETS

Fashion MNIST Dataset:

Fashion MNIST comprises 70,000 grayscale images of fashion items categorized into 10 classes. It serves as a challenging benchmark for image classification tasks. Each image is represented as a flattened array of 28x28 pixels, resulting in a total of 785 features per instance, including labels.

Overhead MNIST Dataset:

The Overhead MNIST dataset comprises 8519 grayscale images of handwritten digits, simulating aerial imagery captured by drones or satellites. Each image is represented as a flattened array of 28x28 pixels, resulting in a total of 785 features per instance, including labels.

WHAT WE DID:

1. Initial MLP Model Construction:

- Created an MLP model using the entire dataset.
- Recorded the accuracy of the model.

2. Utilizing scikit-learn MLP Implementation:

- Employed scikit-learn's MLP implementation to construct another MLP model.
- This provided a standardized approach for comparison.

3. K-Means Clustering and Representative Sampling:

- Conducted K-Means clustering on the dataset.
- Selected one representative sample (nearest to centroid) from each cluster.
- These samples formed the basis for constructing a new MLP model.

4. Experimentation with Different K Values:

- Iterated the process for various values of K in K-Means clustering.
- Evaluated the accuracy of the resulting MLP models for each K value.

5. Augmenting Dataset with Cluster Labels:

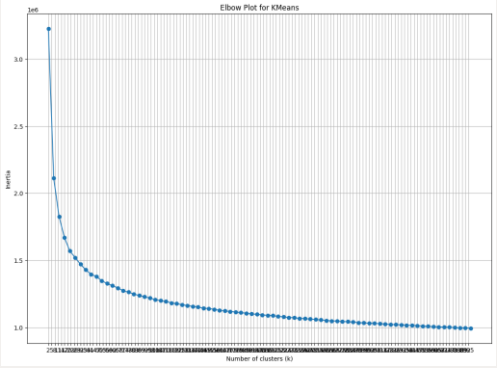
- Assigned cluster labels to the entire dataset.
- Utilized the augmented dataset to construct a new MLP model.

Note: Using Scikit-learn to build an MLP model resulted in very low accuracy in the Overhead MNIST Dataset, so in that case, we tried using Keras to build the model.

RESULTS

Fashion MNIST DATA

- After conducting K-means clustering, we determined the optimal number of clusters (K) to be 125.



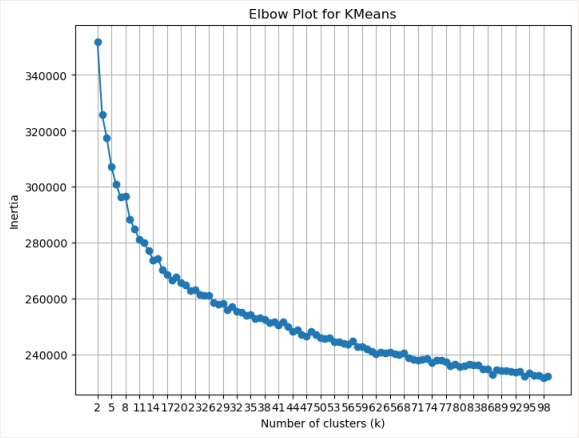
-

Model	Accuracy

MLP using entire data	88.37%	
MLP Using k Representative Samples	**Value of k**	**Accuracy**
	20	61.34%
	50	66.27%
	100	73.22%
	200	75.15%
	500	79.34%
MLP by Extending Cluster Label to Whole Cluster	20	58.43%
	30	62.36%
	50	63.87%
	100	71.37%
MLP by Extending the Label to Only 20% of the Data	10	54.31%
	20	63.25%
	30	59.63%
	50	66.84%

Overhead MNIST Dataset

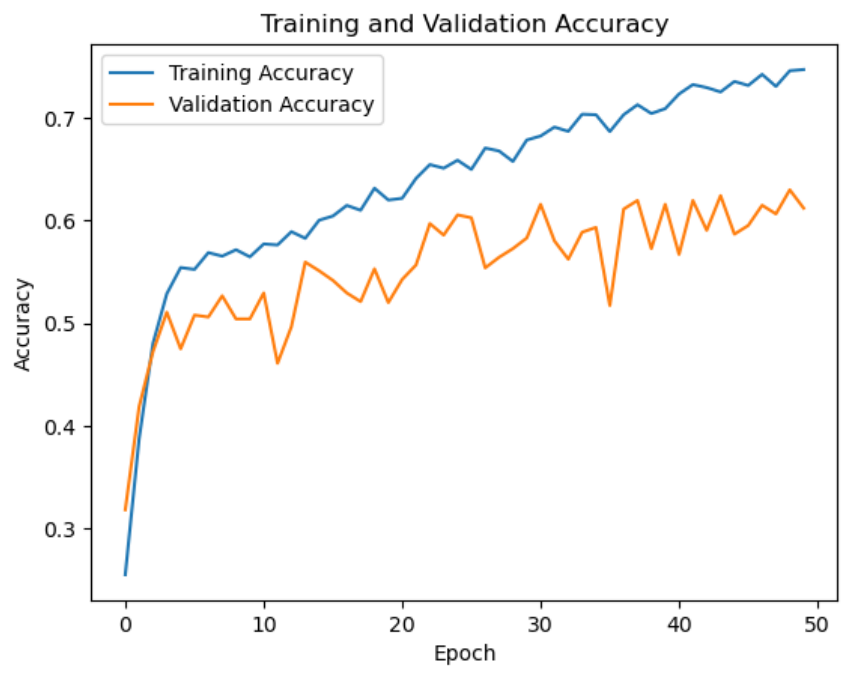
- After conducting K-means clustering, we determined the optimal number of clusters (K) to be 20.



Model	**Precision**
MLP using entire data	0.647451

	Value of k	Precision
MLP Using k Representative Samples	20	0.525921
	50	0.432024
	100	0.415292
	250	0.569046
	500	0.534347
MLP by Extending Cluster Label to Whole Cluster	10	0.718628
	20	0.623166
	50	0.624392
MLP by Extending the Label to Only 20% of the Data	20	0.334227
	30	0.43336
	50	0.203902
	100	0.219275

- The initial accuracy for MLP model was 0.49.
- After switching the optimization algorithm from stochastic gradient descent to Adam, the accuracy increased to 0.6122.



-