Advanced Topics in Machine Learning

Yuri Paglierani

Challenge 1

University of Trieste

1 Introduction

This report presents the results of Challenge 1 in the Advanced Machine Learning course. The challenge involved developing a data analysis pipeline for the Fashion-MNIST dataset, exploring both unsupervised and supervised learning techniques.

2 Source Code

The complete source code, including all implementations and experiments discussed in this report, is available in the following GitHub repository:

`https://github.com/YuriPaglierani/adv-ml-units`

This repository contains the code for the Advanced Topics in Machine Learning practica at the University of Trieste, Spring 2023. Readers are encouraged to explore the repository for more detailed implementations and additional resources related to this project.

3 Dataset

The Fashion-MNIST dataset consists of 60,000 grayscale images of fashion items, divided into 10 categories. Each image is 28x28 pixels, resulting in 784 features.

4 Unsupervised Learning Approach

4.1 Principal Component Analysis (PCA)

First, I applied linear PCA to the dataset and visualized the first two principal components.

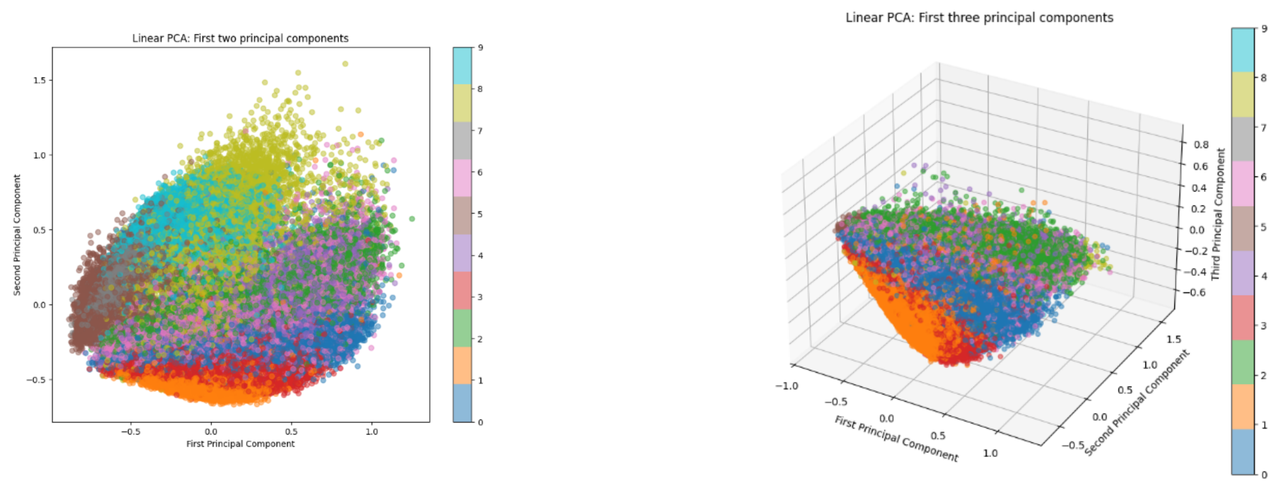


Figure 1: Linear PCA: 2, and 3 principal components

The PCA results showed some separation between classes, but significant overlap remained, indicating the need for more complex approaches.

4.2 Kernel PCA

I then went further, and experimented with the Kernel PCA using both RBF, polynomial, and sigmoid kernels.

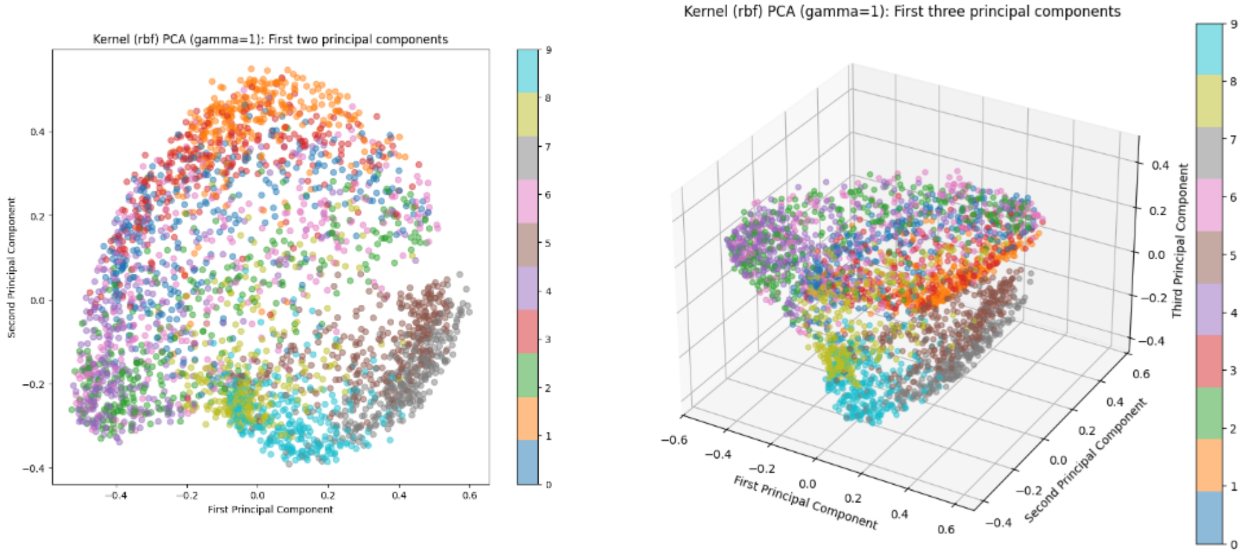


Figure 2: Kernel PCA best model, rbf, gamma=1.0: 2, and 3 principal components

The RBF kernel with gamma=1.0 provided the best separation, capturing non-linear relationships in the data.

From now up to the end, I used a subset of the dataset (1/20), this because the notebook was quite long, and the Kernel PCA is computationally demanding, however main results can still be deduced.

5 Hybrid Unsupervised-Supervised Approach

I then developed a hybrid pipeline combining unsupervised clustering with supervised classification, I wanted to implement a strategy different from the ones proposed but that I tried after dealing with Sankey Diagram, details below.

5.1 Clustering

K-means clustering was applied to the Kernel PCA transformed data, assigning 10 labels to the datapoints.

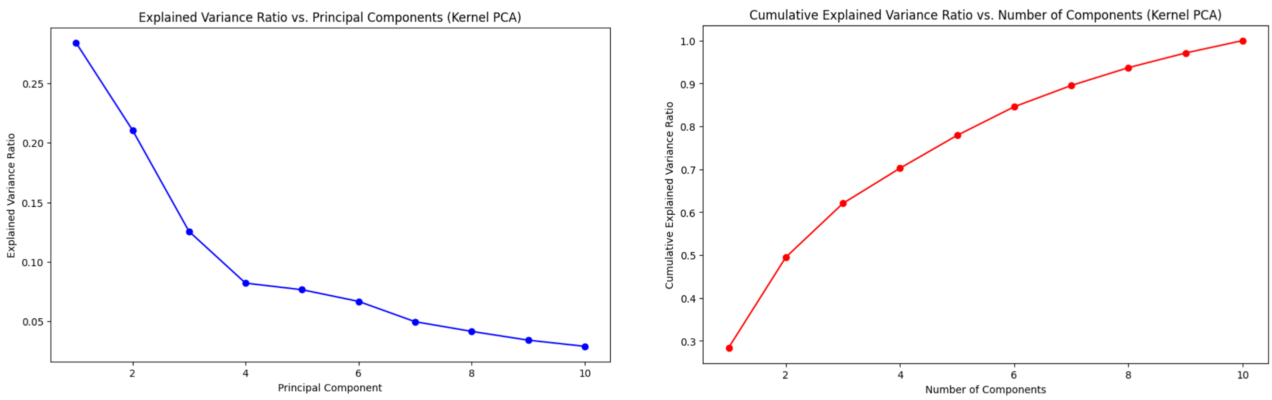
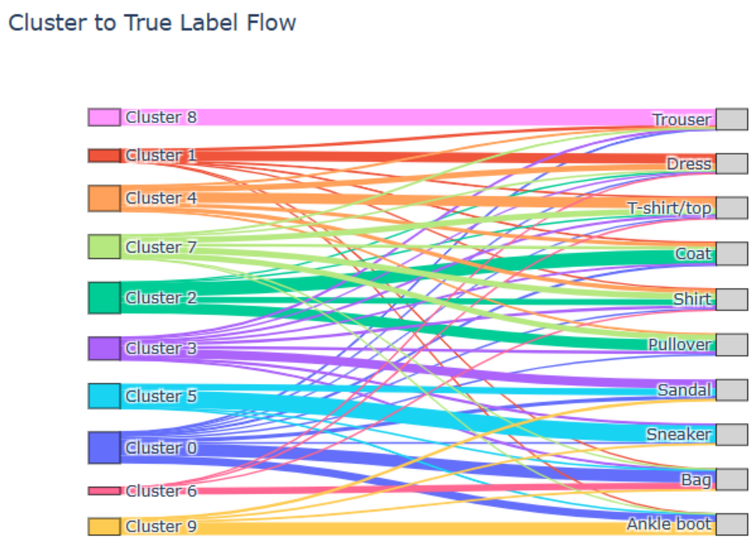


Figure 3: Elbow point, and explained variance plot

We can see that the elbow point is reached for N=4, and for that value we can express more than 70% of the variance in the data.
The Sankey diagram reveals a moderate relationship between clusters and true labels, with an Adjusted Rand Index of 0.362.



5.2 Supervised Classification on Unsupervised Labeling

In this section, I implemented and compared three different supervised classification models: Support Vector Machine (SVM), Fully Connected Neural Networks (FCN), and Convolutional Neural Network (CNN). Each model was trained on the cluster-labeled data obtained from the unsupervised learning phase.

5.2.1 Support Vector Machine (SVM)

I experimented with different kernels (RBF, polynomial, and sigmoid) and gamma values.

- **RBF Kernel:**
  - Best performance: Gamma = 1, Train Accuracy = 99.42%, Validation Accuracy = 97.33%
  - Gamma = 10 showed signs of overfitting (100% train accuracy, 62.33% validation accuracy)
- **Polynomial Kernel:**
  - Consistently high performance across different configurations
  - Best: Train Accuracy = 100%, Validation Accuracy = 95.83%
- **Sigmoid Kernel:**
  - More variable performance
  - Best: Gamma = 0.1, Train Accuracy = 93.71%, Validation Accuracy = 94.17%

The RBF kernel with gamma = 1 showed the best balance between training and validation accuracy, indicating good generalization.

5.2.2 Fully Connected Neural Network (FCN)

I implemented two FCN models with different architectures:

- **FCN1 (smaller network):**

  - Reached 100% training accuracy by epoch 20
  - Best validation accuracy: 85.67% (epochs 16, 19, 20)
  - Shows signs of overfitting in later epochs

- **FCN2 (larger network):**
  - Reached 99.88% training accuracy by epoch 20
  - Best validation accuracy: 86.50% (epoch 18)
  - Slightly better generalization than FCN1, but still shows overfitting

5.2.3 Convolutional Neural Network (CNN)

- Achieved the best performance among neural network models
- Reached 98.54% training accuracy by epoch 20
- Best validation accuracy: 91.00% (epoch 16)
- Showed better generalization compared to FCN models

5.2.4 Comparison and Analysis

1. SVM outperformed neural network models, with the RBF kernel achieving the highest validation accuracy (97.33%).
2. Among neural networks, CNN performed best, likely due to its ability to capture spatial relationships in image data.
3. Both FCN models showed signs of overfitting, with training accuracy approaching 100% while validation accuracy plateaued.
4. The CNN demonstrated better generalization, with a smaller gap between training and validation accuracies.

These results suggest that for this specific task of classifying cluster-labeled Fashion-MNIST data, the SVM with RBF kernel was most effective. However, the CNN's performance indicates that with further optimization (e.g., regularization, data augmentation), neural network approaches could potentially match or exceed SVM performance.

The high accuracies achieved by these models on cluster-labeled data suggest that the unsupervised clustering captured meaningful structure in the dataset. However, it's important to note that these accuracies are with respect to the cluster labels, not the original Fashion-MNIST labels, which explains the unusually high performance compared to typical results on this dataset.

5.2.5 Performance on Test Set with Real Labels

To evaluate the models' performance on the original Fashion-MNIST labels, I employed a probabilistic mapping from cluster predictions to real labels. For each cluster prediction, we sampled from the distribution of true labels within that cluster, as observed in the training data. This approach allows us to assess how well our unsupervised-to-supervised pipeline generalizes to the original classification task.

Table 1 presents the classification reports for each model on the test set, using this probabilistic mapping to real labels.

Key observations from the test set performance:

1. Overall accuracy: all the models has approximately the same accuracy, 44% for SVM, 41 for the CNN, and 42% for the FC neural networks.

Class	**SVM**	**FCN1**	**FCN2**	**CNN**
T-shirt/top	0.30	0.29	0.31	0.30
Trouser	0.83	0.76	0.76	0.79
Pullover	0.29	0.27	0.26	0.24
Dress	0.49	0.48	0.47	0.45
Coat	0.32	0.30	0.31	0.31
Sandal	0.29	0.28	0.27	0.28
Shirt	0.17	0.19	0.16	0.19
Sneaker	0.59	0.60	0.60	0.59
Bag	0.54	0.48	0.50	0.48
Ankle boot	0.55	0.53	0.53	0.54
Accuracy	0.44	0.42	0.42	0.41

Table 1: Accuracy for each model on the test set using probabilistic mapping to real labels

2. Class-wise performance:
    - Trouser is consistently the best-classified category across all models, with accuracy ranging from 0.76 to 0.83.
    - Shirt is the most challenging category, with accuracy ranging from 0.16 to 0.19 across all models.
    - Categories like Pullover, T-shirt/top, and Sandal show relatively low accuracy (around 0.3) across all models.

These results indicate that while our unsupervised-to-supervised pipeline captured some meaningful structure in the data, there is a significant drop in performance when mapping to the original Fashion-MNIST labels. The consistent difficulty across all models in distinguishing certain categories (e.g., Shirt, Pullover) suggests that these classes may have been mixed in the initial clustering phase.

The relatively low accuracies (41-44%) compared to typical supervised learning results on Fashion-MNIST (which can exceed 90%) highlight the challenges of this unsupervised-to-supervised approach. However, the fact that all models perform significantly above random guessing (10% for a 10-class problem) indicates that the pipeline has indeed learned useful features and patterns from the unlabeled data.

6 Fully Supervised Approach

I repeated the classification using true labels for comparison.

6.1 Model Performance

Table 2 compares the test accuracies of the hybrid and fully supervised approaches.

Model	Hybrid Accuracy	Fully Supervised Accuracy
SVM	44%	81.19%
FCN1	42%	69.42%
FCN2	42%	69.57%
CNN	41%	72.57%

Table 2: Model Performance Comparison

6.2 Class-wise Performance

The fully supervised approach significantly improved classification across all classes, with some consistently challenging classes (e.g., Shirt) and some easily distinguishable ones (e.g., Trouser).

7 Discussion

7.1 Unsupervised Learning Insights

The PCA and Kernel PCA analyses revealed the complex, non-linear nature of the Fashion-MNIST dataset. The unsupervised clustering captured some meaningful structure, as evidenced by the moderate Adjusted Rand Index.

7.2 Hybrid vs. Fully Supervised Approach

The hybrid approach demonstrated the potential of using unsupervised methods to create preliminary classification systems, achieving above-random performance. However, the substantial performance gap with the fully supervised approach (24-33% improvement) highlights the value of labeled data in this task.

7.3 Model Comparison

Surprisingly, SVM outperformed neural network models in the fully supervised setting, achieving 81.19% accuracy. This suggests that the decision boundaries in the Fashion-MNIST dataset might be effectively captured by the SVM's kernel trick, howerver if we don't want to compare with SVMs, we can use the full dataset, and improve drastically (especially in the ambiguous classes).

8 Conclusion

This challenge demonstrated a toy model to deep dive in the topic of Kernels, and Neural Networks. While the fully supervised approach significantly outperformed the hybrid pipeline, the unsupervised methods provided valuable insights into the data structure. The surprising efficacy of SVM in this image classification task warrants further investigation and could inform future model designs.