Feasibility of automating image classification of clothing items using machine learning

Enrico Ancilotto

Imperial College London

Introduction

MNIST-fashion is a dataset containing labeled images of clothing items belonging to 10 different categories. We will study the feasibility of using both supervised and unsupervised methods to automatically classify the pictures.

Figure 1: Selected pictures from the dataset

Unsupervised Learning: *k*-Means

Unsupervised learning is performed by using the *k*-means clustering algorithm, which splits the data into *k* clusters trying to have clusters which are as small and as far from each other as possible. Each image is treated as a 784-dimensional vector, where each dimension represents the brightness value of a single pixel. We set $k = 10$ to try to split the data into the 10 classes.

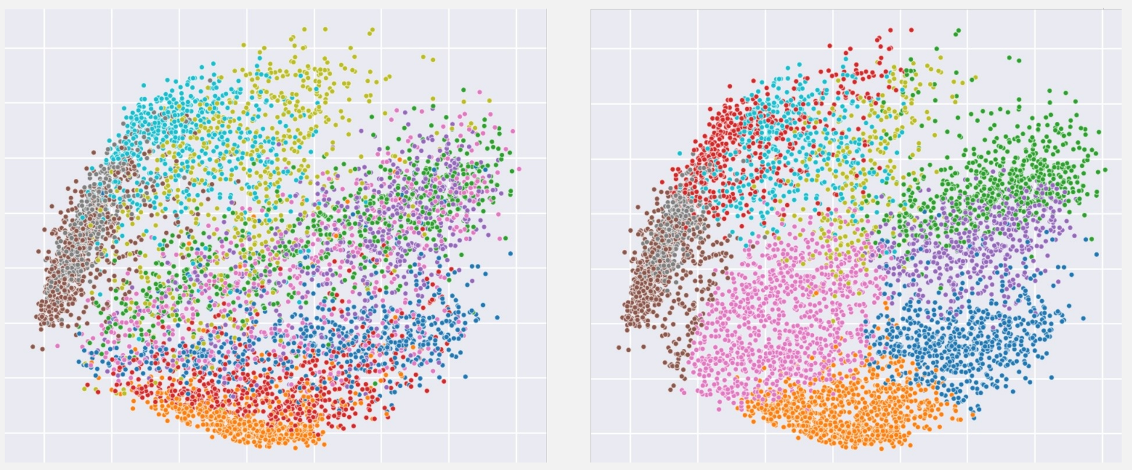

Figure 2: True labeling of the data (left) and labeling predicted by *k*-means (right). Each image is projected into 2D by using PCA to determine the dimensions with highest variance

Supervised Learning: MLP

The first supervised method we used was that of a neural network with a traditional fully connected architecture. The model was composed of three hidden layers, each with 100 nodes. The network was then trained for 90 epochs.

Supervised Learning: CNN

The second approach to supervised learning was to use a convolutional neural network. They have proved to be superior to traditional neural networks at image recognition tasks [1]. The network is comprised of 2 alternating convolutional and subsampling layers, followed by 3 fully connected layers.

Supervised Learning: Improved CNN

Since the initial CNN architecture didn't show the expected performance, a new architecture was used, inspired by Google's Inception-ResNet [2]. The new model was the only one to be able to distinguish between all 10 classes, confirming the superiority of convolutional methods for image classification tasks.

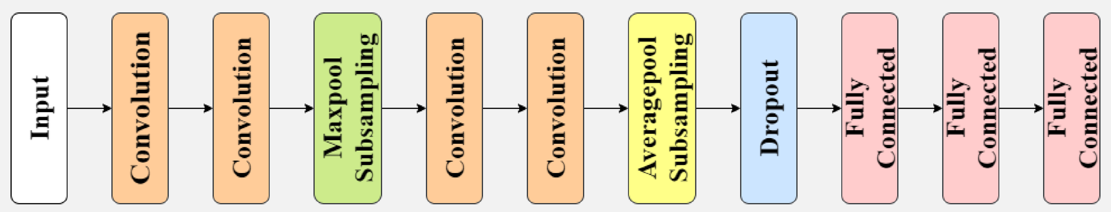

Figure 3: Schematic view of the architecture of the improved CNN

Results

Method	Precision (%)	Recall (%)
k-Means	58	53
MLP	63	72
CNN	42	55
CNN Improved	91	91

Unsupervised Learning: Findings

As can be seen in figure 2, *k*-means has done a good job at dividing the images into clusters, however it failed to get as good of a precision as could be expected given that. This can be explained by the clustering revealing a different ground truth: the clothing items are divided by picture similarity rather than use. So, for example, a light t-shirt and shirt will be more similar than a light shirt and a dark one.

In any case the labeling is still quite consistent with the correct one, getting only one class completely wrong.

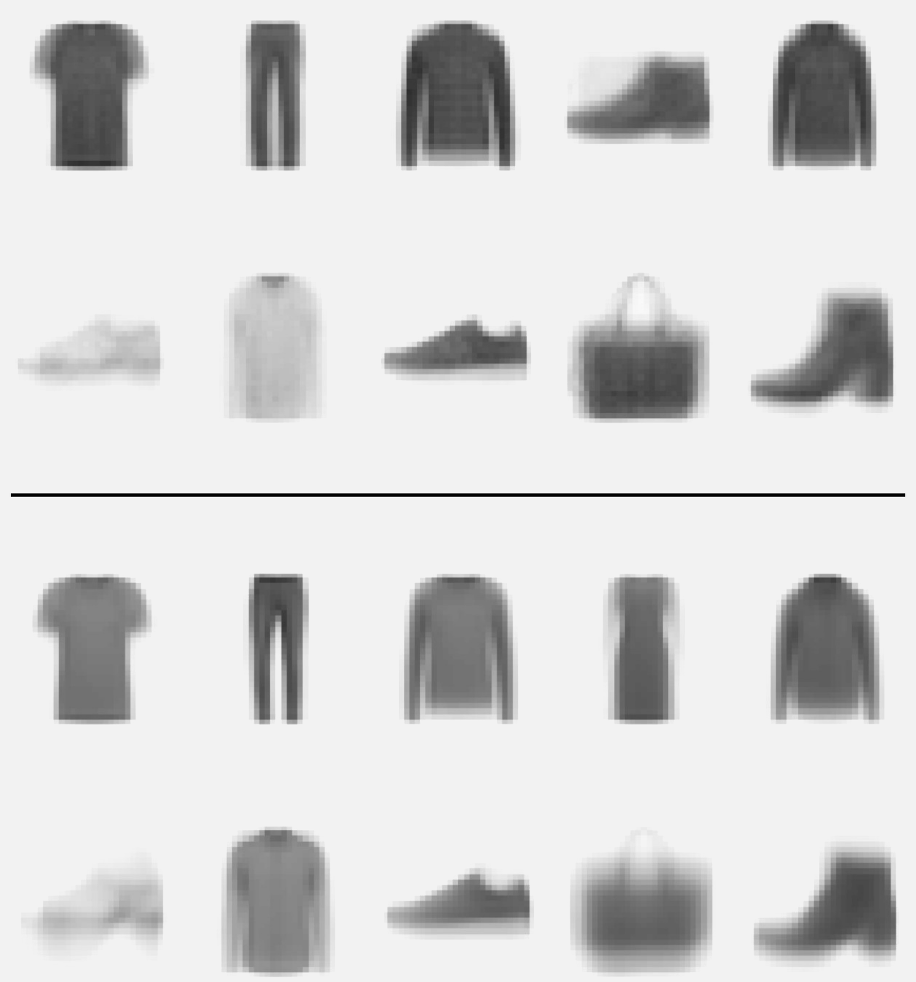

Figure 4: Centroids of the clusters obtained from the *k*-means clustering algorithm (top) and mean of the true clothing classes (bottom)

Supervised Learning: Findings

From the confusion matrices (figure 5), a very clear limit on the ability to distinguish between the visual features for the MLP and CNN can be seen: the network never attempts to distinguish between classes for which it is unsure, meaning it will never predict certain classes.

To overcome this, imitating already existing architectures can give extreme improvement in performance, without the cost associated with their development.

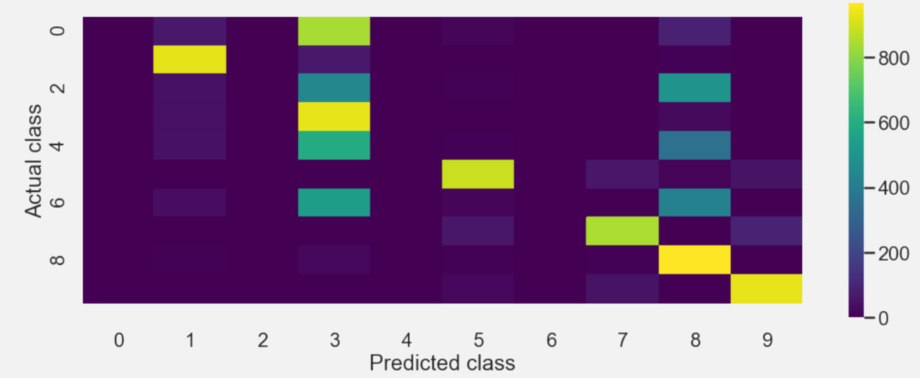

Figure 5: Confusion matrix for the CNN architecture

Conclusions

Though less accurate, the relative simplicity of unsupervised learning models means that they can be used for exploratory data analysis and to serve as a baseline.

On the other hand, thanks to the free availability of supervised ML models, the development cost of sufficiently accurate image recognition tools is low enough to justify using it to automate existing pipelines.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton.
Imagenet classification with deep convolutional neural networks.
In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[2] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi.
Inception-v4, inception-resnet and the impact of residual connections on learning.
In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.