DA & DM PROJECT

For this project we implemented a Neural Network with one hidden layer composed by 30 neurons in order to predict the type/category of clothes given an image representing the clothes item.

Dataset

Fashion-MNIST is a dataset of Zalando’s article images consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28×28 grayscale image, associated with a label from 10 classes.

Fashion-MNIST is intended to serve as a direct drop-in replacement of the original MNIST dataset for benchmarking machine learning algorithms. The 10 classes of the labels are: T-shirt/Top, Trousers, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag and Ankle Boot.

<https://github.com/zalandoresearch/fashion-mnist>

Neural Network

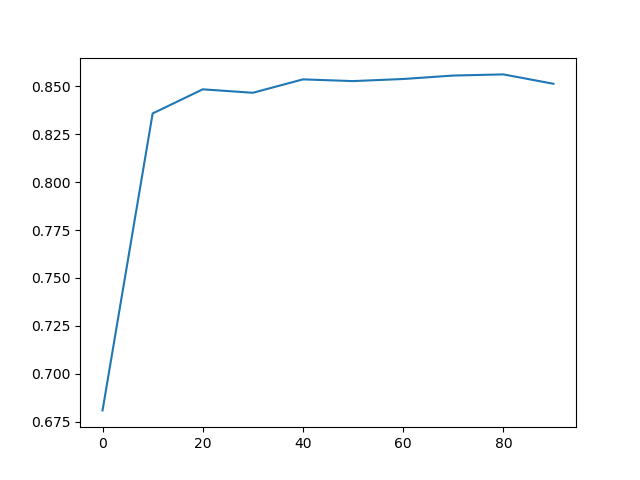
In order to observe and discuss the differences caused by the optimizations techniques, we initially started developing a neural network with one hidden layer and 30 neurons which has the Sigmoid Function as Activation Function both for the hidden layer and for the output layer, composed by ten neurons, and the Mean Square Error Function as Cost Function. 

Figure 1 - mini-batch of size 10, learning rate equals to 0.5

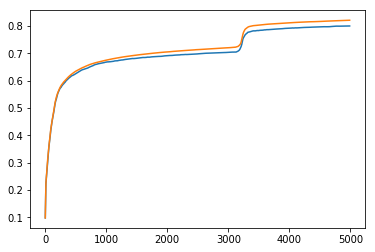


Figure 2- batch gradient discend, learning rate equals to 1.5

Then we tried to improve this first Neural Network by changing the weights inizialization and the cost function using the Cross Entropy Function and implementing the L2 regularization.

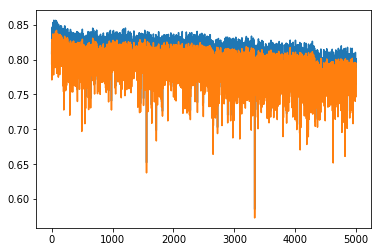
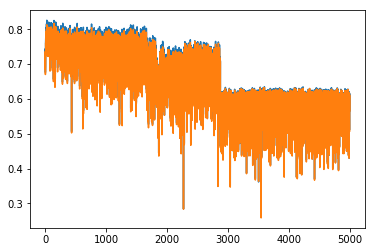


Figure 3-mini batch of size 5, learning rate equals Figure 4- mini-batch of size 5, learning rate equals to 0.5,

to 1.5, lambda equals to 0.5 lambda equals to 2.5

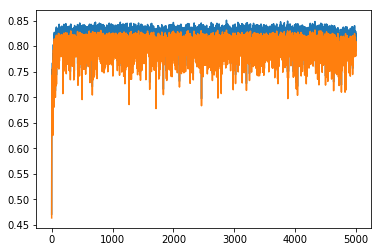
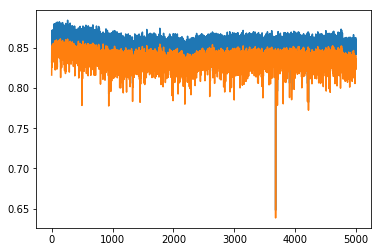


Figura 1-mini-batch of size 1, learning rate equals to 0.05, lambda equals to 3.5

Figure - mini batch of size 1, learning rate equals to 0.5,

lambda equals to 0.05

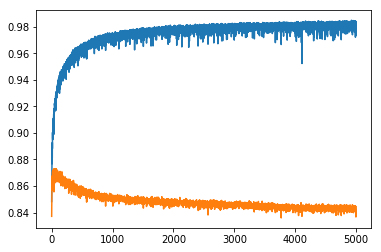
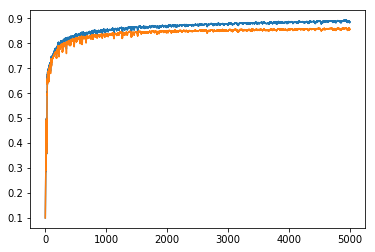


Figure 6-batch gradient discent, learning rate equals to 1.5, lambda equals to 0.5

Figure 7- batch gradient discent, learning rate equals to 0.5, lambda equals to 0.05

Finally we introduce the early stopping technique to avoid overfitting and the momentum to speed up the training phase.

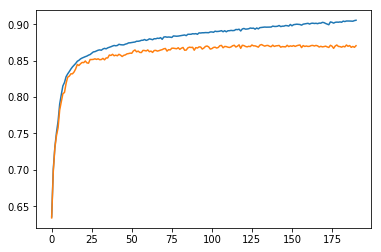
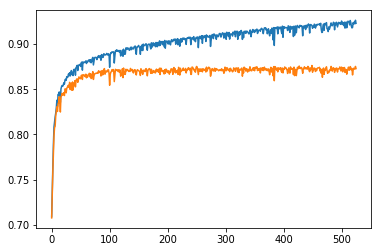


Figure 8-mini-batch of size 1000, learning rate equals to 0.05,

lambda equals to 0.05, stop equals to 100 without momentum

Figure 9- mini-batch of size 1000, learning rate equals to 0.05,

lambda equals to 0.05, stop equals to 30

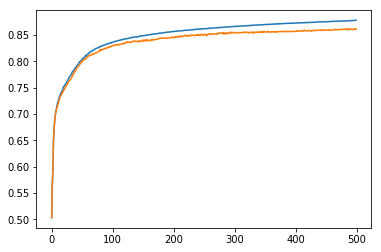
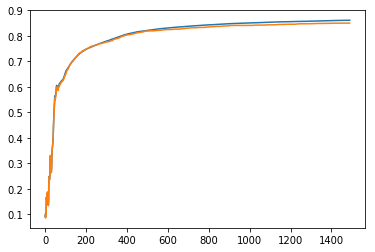


Figura 2-batch gradient discent, learning rate equals to 0.05, lambda equals to 0.05, stop equals to 100

Figure 10- batch gradient discent, learning rate equals to 0.05,

lambda equals to 0.05, stop equals to 30