Machine Learning - 89511

Assignment 3 - Solution

Boaz Ardel - 203642806

May 2018

Department of Computer Science

Bar-Ilan University

´LaTeX

Contents

Chapter 1

Theoretical Part

1.1 NN Training Hyper-Parameters

A Network design:

As required we have 2-layer NN for training on MNIST clothing dataset.

The best Performance produced by ReLU non-linear function for first layer, in addition ReLU is not sensitive for big values thus not getting saturated, comparing to sigmoid/tanh functions. For second Layer the general softmax function has been used, influenced at most on the η because of overflowing at some point.

B Hyper-Parameters:

Best practice were H = 128 for hidden Layer, more than this was "over-fitting" and did not increased performance. The number of epochs that yields the optimal outcome comparing to computational time is 24, more than this yields less than 0.1Best learning rate of $\eta = 0.0005$ gave good congestion, Softmax did not overflowed with this value.
mini-batch did not worked well, computation time increased badly, but no better results appeared.

C Results:

I had 89% success rate for the hyper-parameters discussed above, a bit of weight tuning was needed, as a result of overflowing.

Chapter 2

Practical Part

The code is in attached files as instructed.