Itay Hassid 209127596 Aviv Shisman 206558157

Exercise 1 - report

Intro - about the model

pre train

We start of by loading the different data sets(train/valid/test) and creating the weights with the proper dimensions the first one (hiddenLayer , inputLayer) and the second (outputLayer , hiddenLayer) both given randomly uniform values that we will change later during training.
Notice that the hidden layer is one of the hyper parameters and the input and output layers are both constant with the values of 3072 and 10.
To conclude we have 3 layers : input(with 3072 neurons), hidden with changing number of neurons depends on hiddenLayer variable , output (10 neurons) corresponding to the 10 classes.

Training

in training we go over the training set “epoch” times (hyper param) and in each one we shuffle the train set to simulate the SGD algorithm randomness.
In each epoch we start of by running the input through the layers using the activation function(simoid/relu) and softmax for output → feedForward function.
After that we calculate the gradients in the back propagation stage using the derivatives we calculated before the exercise (each activation function has its own derivatives) , Then we use the gradients to update the weights.

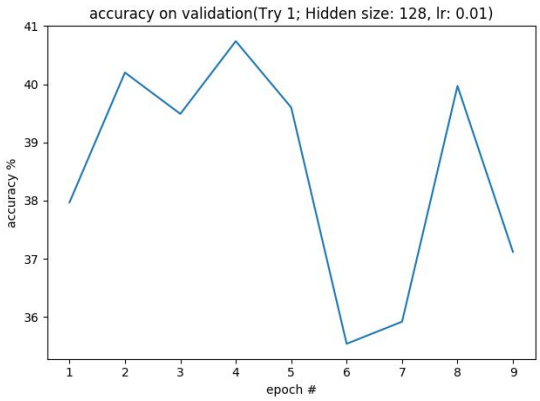
Every 3\4 epochs (changes in versions) we check the results of our model on the validation set for accuracy and loss measuring while avoid overfit to train set.
After “epoch” times training is done, the model weights are tuned and we are ready , so we write our predictions on the test set.

Notes

1. In the results below some of the versions used sigmoid and some relu we sent the code and the output.txt of the best one.
2. The hidden layer size and learning rate is also changing between versions.
3. Some of the versions we used dropout.
4. Because we did validation every ¾ epochs it will be visible in the graphs(in the first versions we did every epoch or 2), you can see the results below.
5. The graphs below are the accuracy on validation set

Itay Hassid 209127596 Aviv Shisman 206558157

Results

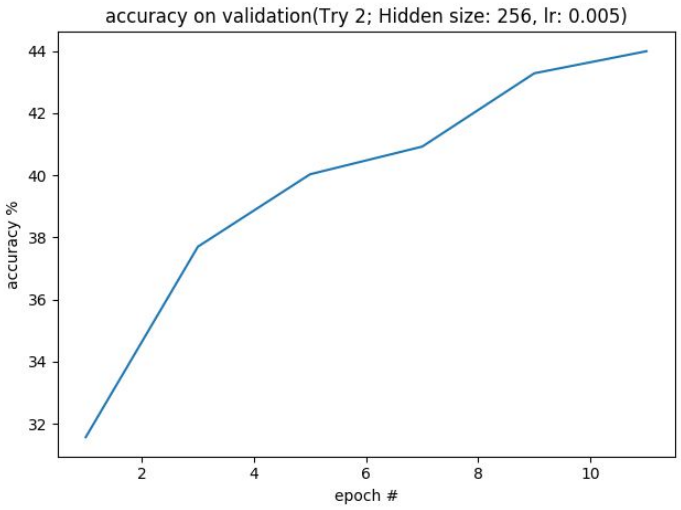


Version 1

The first attempt wasn't very successful,the learning rate we used was too high ,so we quickly overfitted to the train set.

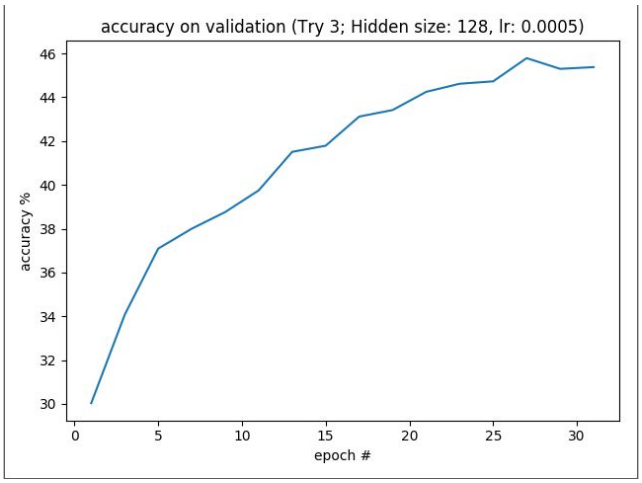
Version 2

We improved the results by lowering the learning rate helped us avoid overfitting to train set and get better results on valid set.
We also raised the hidden layer size, which resulted in low train speed but better accuracy.



Version 3

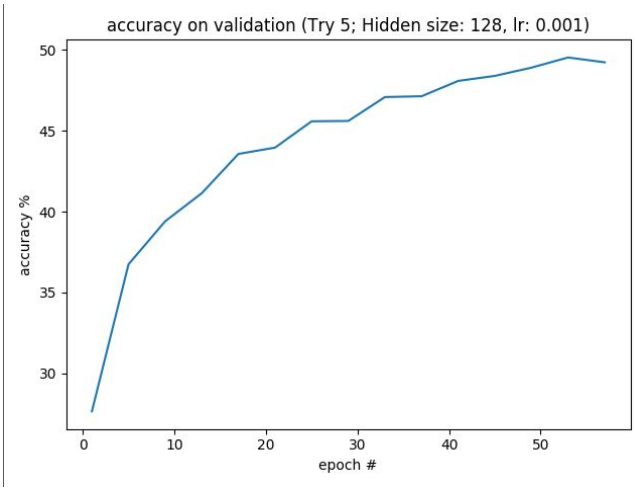
We improved the efficiency of feed_forward function and the loading of the data sets , we also increased the number of epochs and lowered the learning rate which brought us a good slow ascension over time.

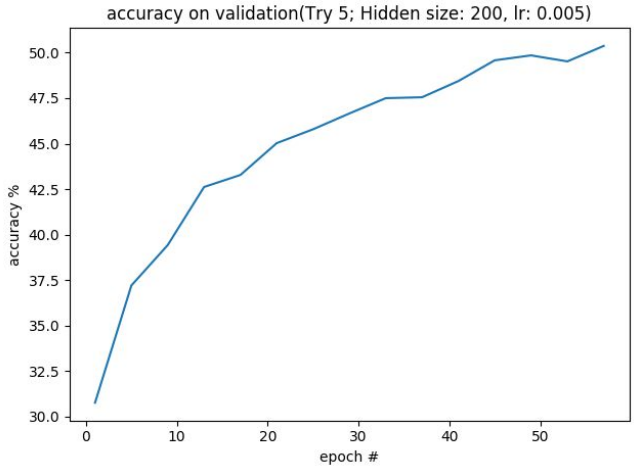


Itay Hassid 209127596 Aviv Shisman 206558157

Versions 4 and 5

We improved the softmax function to be more efficient,
We also fixed the derivatives to be more accurate (in the back prop function) ,we changed the learning rate to fit to the new changes and increased the number of epochs → all of that brought us better results.





Version 6

In this version we added dropout.
The dropout was on the hidden layer and we tried a few dropout parameters: 0.5, 0.6 and 0.8.
At this point we also tried higher hidden size layers but we could not get higher results the 50%.