Deep Learning in Data Science, DD2424

Short report on lab assignment 1

Bonus points

Anna Sánchez Espunyes

April 12, 2021

1 Main objectives and scope of the assignment

In this assignment you will train and test a two layer network with multiple outputs to classify images from the CIFAR-10 dataset. You will train the network using mini-batch gradient descent applied to a cost function that computes the cross-entropy loss of the classifier applied to the labelled training data and an L2 regularization term on the weight matrix.

2 Results and discussion

2.1 Bonus 1: Improvements

1. Increasing the number of hidden nodes: The first improvement consists on increasing the number of hidden nodes and observe if it leads to better validation accuracy. In the results (see Figure 1) we can see the improvement in validation accuracy when increasing the number of hidden nodes. Each curve represents different amount of regularization. As more hidden nodes, we need to use more regularization in order to not overfit. Another aspect worth noticing is that a plateau in accuracy is reached when using 750 nodes. As it is computationally expensive to add more hidden nodes, it is not worth it to have more than 750 hidden nodes.

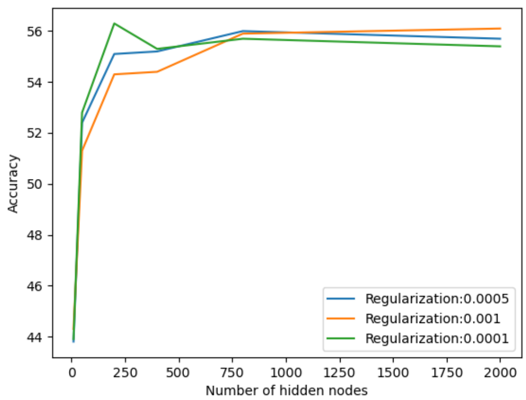


Figure 1: Accuracy in the validation set for different number of hidden nodes

2. Adding dropout: Dropout helps us with regularization. It allows training deep networks and makes them less prone to over-fitting. During the training phase, it sets a certain amount of neurons to 0. Intuitively, we are making the network learn different types of neuron subsets. As dropout is a useful

technique when using a lot of hidden nodes. For the experiments, 750 nodes have been used to perform the training in 3 cycles. The evaluation has been done comparing the validation accuracy when using dropout and when not using dropout(see Figure 2). As not very deep networks are being used, dropout does not provide a better performance.

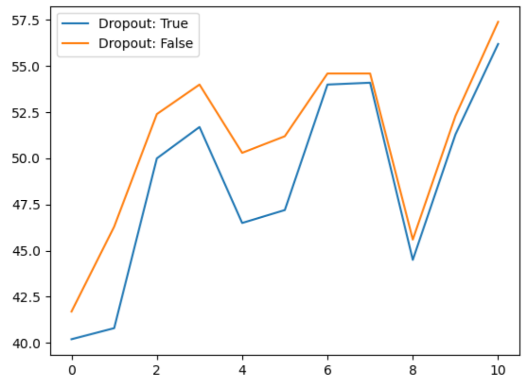


Figure 2: Accuracy in the validation set over different iterations

3. Adding jitter Adding jitter also helps with regularization. If the available data set is too small, the network may memorize the samples. By adding Gaussian noise when training it is more unlikely to over-fit. The mean of the noise added is 0 and the variance 0.1. In order to experiment with jitter, a network of 750 hidden nodes has been used, without dropout. The results (see Figure) show that using Jitter increases the validation accuracy.

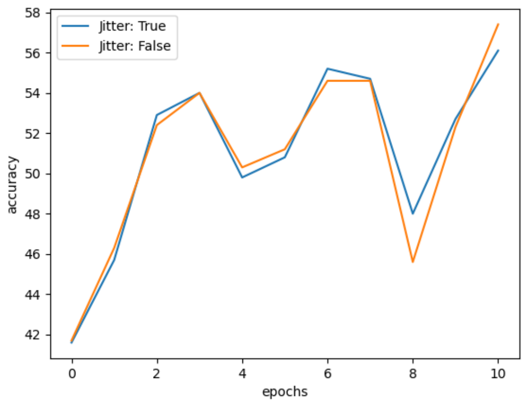


Figure 3: Accuracy in the validation set over different iterations

Best model: The final hyper-parameters that have given best validation accuracy are 750 hidden nodes, regularization of 0.001, dropout, jitter, 3 cycles, batch size of 90, minimum learning rate 1e-5 of and maximum learning rate of 1e-1. The test accuracy achieved using these hyper-parameters is 56.1. The improvement that brought the largest gain is increasing the number of hidden nodes. Just in comparison, the maximum test accuracy that I obtained in the basic part was 51.47.

2.2 Bonus 2: Eta calculation

In the basic part of the assignment the cyclic learning rate was implemented. This technique is useful in order to find a suitable learning rate without a wide experimentation. In this method, the learning rate is increased and decreased between two boundaries. With this implementation a better accuracy was achieved, but we assumed some boundaries. In this exercise the boundaries will no longer be assumed, in fact a test will be performed in order to find the right parameters.

The experiment that will be performed is called "LR range test". It basically consists of setting the learning rate into a very small number and increasing it linearly to the maximum value during a short run. By plotting the accuracy vs the learning rate, it can be seen when the accuracy starts increasing (this is the minimum learning rate) and when the accuracy reaches a plateau (this is the maximum learning rate).

The results of the experiment can be seen in Figure 4. The accuracy starts increasing when the learning rate is approximately 1e-5, which will be set as the minimum learning rate. On the other hand, the accuracy reaches a plateau when the learning rate is approximately 0.02, which will be set as the maximum learning rate. After setting these new parameters, we can see that there is no improvement in the loss/cost/accuracy (see Figure 5). Nevertheless, these technique has provided a reliable methodology to find these values.

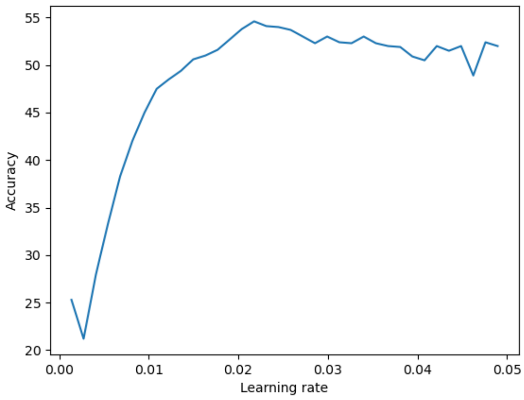
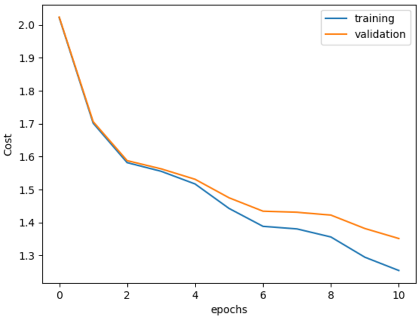
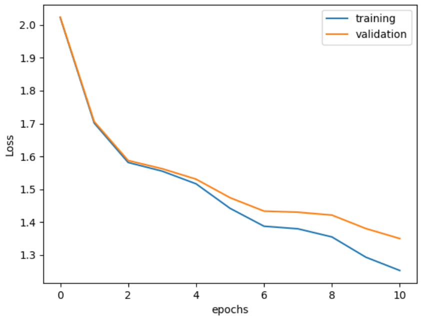


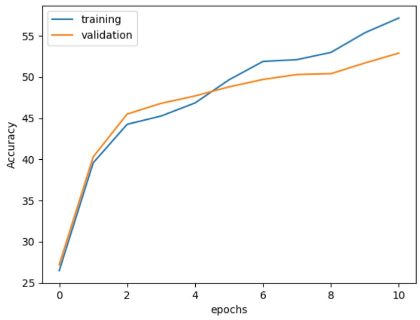
Figure 4: Accuracy in different learning rates (performing the LR range test)



(a) Cost



(b) Loss



(c) Accuracy

Figure 5: Loss, cost and accuracy with the new eta min and eta max