Kungl Tekniska Högskolan

MSc Machine Learning

DD2424 Deep Learning in Data Science

Assignment 2 Bonus: Image classification with a 2-layer network

---

Author:
Alexandros Ferles
ferles@kth.se

Professor:
Josephine Sullivan

Contents

1 Exercise 1: Optimize the performance of the network

> *Try at least 3 improvements (beyond using the all the training data) to help bump up performance and report your results.*

The following optimizations have been performed:

1. **Training using all the available data, except 1000 images used for validation, for a larger number of epochs and selecting the best model based on validation-set accuracy performance**:

   In this setting, we make use of two pair of values for the learning rate η and the amount of regularization λ, based on the results derived from the compulsory part. The number of epochs is extended in order to train for more update steps, and the model that achieved the best validation set accuracy performance during training is selected. The results in the test set are presented in the following table:

Setting #	λ	η	Test set accuracy
1	0.02878809988519304	0.001	**51.53%**
2	0.018920249916784752	0.001	**51.94%**
3	0.018920249916784752	0.0001	**49.94%**

   The cross entropy loss evolution on the training and validation set is presented in the following plot:

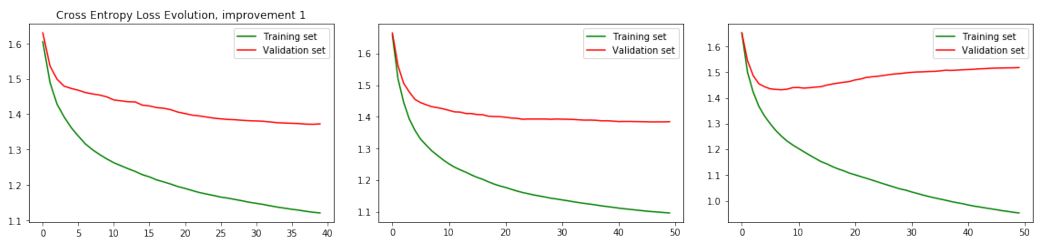

Figure 1: Cross entropy loss evolution for setting 1 (left), setting 2 (center) and setting 3(right)

2. **Switching to He initialization and observing the effect it has on training**. Instead of random initialization, the variance of the normal distribution is dependent of the size of the hidden and output layer, according to He initialization. The values for η and λ in this setting and the following settings (except stating something else), are the pair $(\eta, \lambda) = (0.01892024991678475, 0.001)$ which had better test set accuracy performance compared to the other 2 options tested.

   The evolution of cost when He initialization is applied on the weights is the following:

   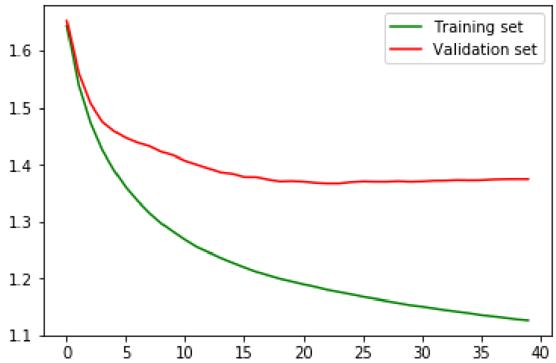

   Figure 2: Cross entropy loss evolution with He initialization on the weights

   The test set accuracy performance of He initialization is 51.44%, which is a better performance compared to the training with random initialization on the weights and also training for 30 epochs. We can also observe a slightly higher drop in the validation loss, which can also explain the fact that the accuracy perormance was better at this setting.

3. **Increasing the number of nodes in the hidden layer**. The number of nodes used in the hidden layer of the network grows up to 100, instead of 50.

Figure 3: Cross entropy loss evolution with 100 nodes in the hiddenlayer

The accuracy of this setting method is 53.48%, which is also the best performance achieved through an improvement tried a exercise 1.

4. **Dropping the learning rate in half every 5 epochs**

The cross-entropy loss evolution of this setting is the following:

Figure 4: Cross entropy loss evolution with annealing the learning rate

The loss evolution when using learning rate annealing is 51.34%:

5. **Augmenting the training data on the fly, applying random transformations to them before doing the forward and backward pass**.

   During this setting, at each epoch a random jittering is applied to the training data, including random rotations, shifts and a possible horizontal flip. After these transformation are applied, with probabillity 50%, random noise is applied to each image.

   Some of the augmented images, compared to the original ones can be seen below:

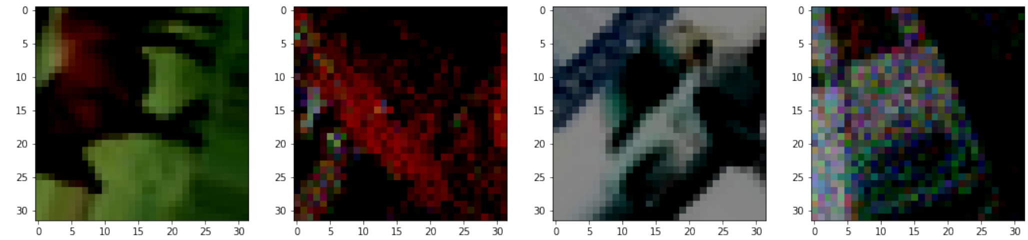

Figure 5: Samples of augmented images

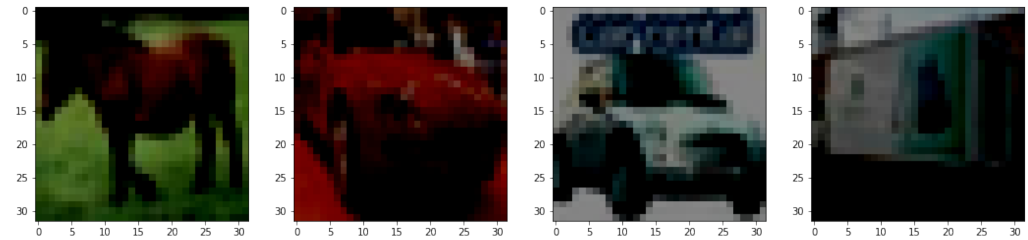

Figure 6: Corresponding original images

The cross-entropy training and validation set loss evolution for each case is shown for each case respectively in the following plot:

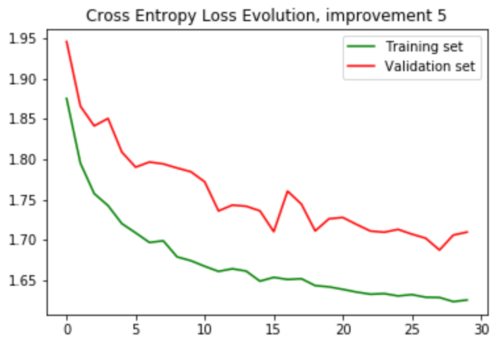

Figure 7: Cross evolution loss with image jittering

The test-set accuracy performance of this setting is only 40.35%:

2 Exercise 2: Train network using different activation to ReLU

Use one of the other activation functions described in the lecture notes, for several sensible parameter settings.

In this part, ReLu is replaced by the Leaky ReLU activation function:

$$Leaky\_ReLU(x) = \begin{cases} a * x, & x < 0 \\ x, & x > 0 \end{cases}$$

where in this case $a = 0.01$

Several tries are performed, including some of the original settings along with combinations of the Leaky ReLU with optimizations performed at the first part of this assignment:

1. Leaky ReLU with $\eta = 0.018920249916784752$, $\lambda = 0.001$
   A test set accuracy of 51.47% is achieved, while the cross-entropy loss evolution is the following:

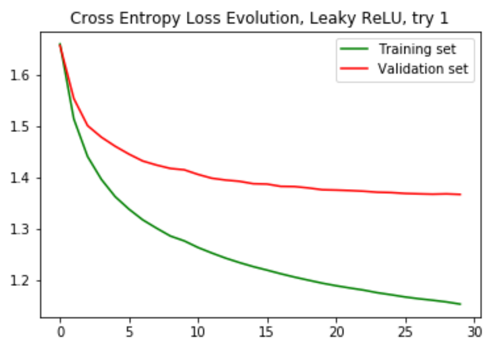

Figure 8: Cross entropy loss evolution for Leaky ReLU, first try

2. Leaky ReLU with $\eta = 0.018920249916784752$, $\lambda = 0.0001$
A test set accuracy of 50.28% is achieved, while the cross-entropy loss evolution is the following:

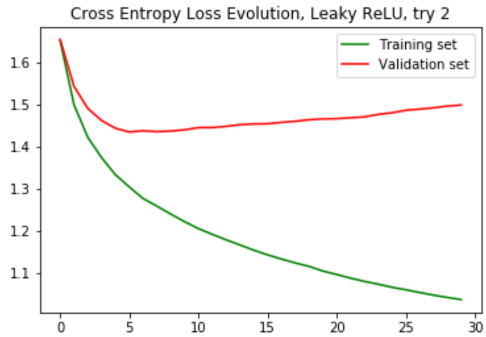

Figure 9: Cross entropy loss evolution for Leaky ReLU, second try

3. Leaky ReLU with $\eta = 0.028788099885193 04$, $\lambda = 0.001$
A test set accuracy of 51.86% is achieved, while the cross-entropy loss evolution is the following:

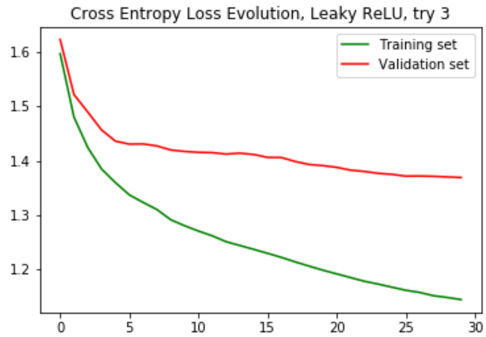

Figure 10: Cross entropy loss evolution for Leaky ReLU, third try

4. Leaky ReLU with learning rate annealing
   A test set accuracy of 51.24% is achieved, while the cross-entropy loss evolution is the following:

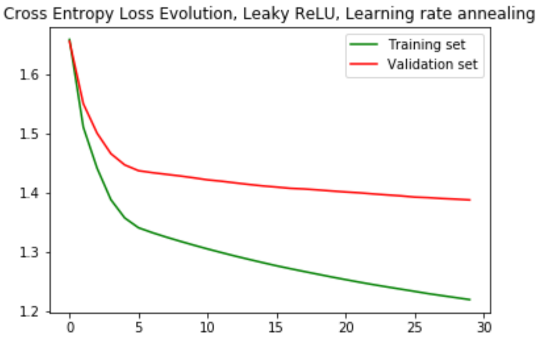

Figure 11: Cross entropy loss evolution for Leaky ReLU combined with annealing the learning rate

5. Leaky ReLU with He initialization
   A test set accuracy of 51.53% is achieved, while the cross-entropy loss evolution is the following:

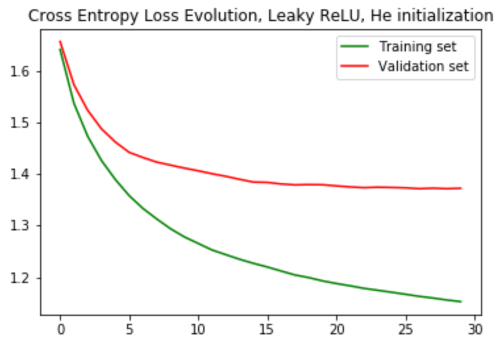

Figure 12: Cross entropy loss evolution for Leaky ReLU combined with He initialization

6. Leaky ReLU with increasing the number of hidden nodes
   A test set accuracy of 53.62% (overall best throughout the assignment) is achieved, while the cross-entropy loss evolution is the following:

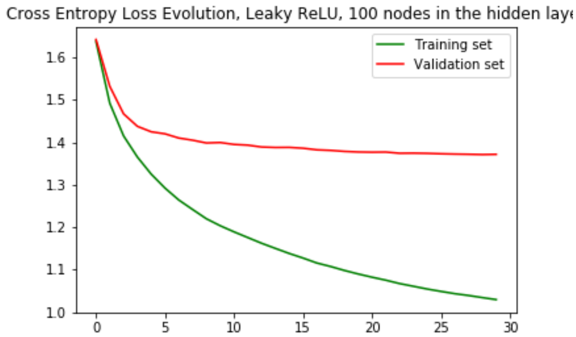

Figure 13: Cross entropy loss evolution for Leaky ReLU with 100 nodes in the hidden layer

7. Leaky ReLU with He initialization and increased number of nodes in the hidden layer.
   A test set accuracy of 53.05% is achieved, while the cross-entropy loss evolution is the following:

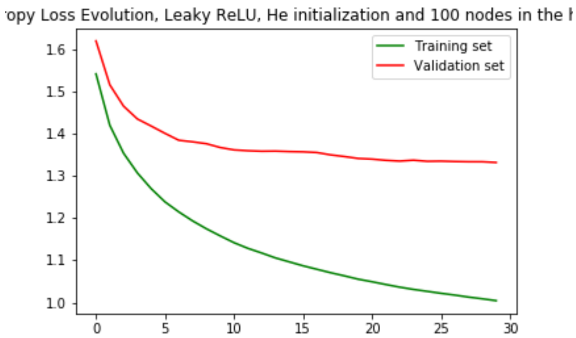

Figure 14: Cross entropy loss evolution for Leaky ReLU with random jittering on the fly

To recapitulate, of the improvements tried throughout this assignment, increasing the number of nodes in the hidden layers proved to be the best possible, managing to achieve performance greater than 53.5%, while every other setting could not achieve even 52% accuracy performance on the test set. Moreover, the use of Leaky ReLU in most cases proved to be an improvement in comparison with settings that made use of the classic ReLU as an activation function. This phenomenon can be attributed to the fact that instead of completely removing the negative activations, it takes them into account allowing them to provide a small contribution to the predictions. This in fact is a very important observation, which allows for more possible explorations of different activation functions with similar behavior, such as the ELU and the Swish activation functions.