IT3105 Project module 5

Deep Learning for Image Classification

Iver Jordal

Five different artificial neural networks

I've experimented with a few different combinations of number of hidden layers, sizes of hidden layers and activation functions in order to get good results in the training set, the validation set and the demo set. I will now present each combination.

Each run was stopped after 200 epochs. This has proved to be enough to get a good idea of how well the network configuration performs.

In order to get reproducible results, a pseudo random number generator (PRNG) is used for random values. One of the parameters of the training module is the seed for the PRNG. When training 20 networks, a different seed (in the range [1, 20]) is used for each of the 20 networks.

Configurations

Network id	1	2	3	4	5
Number of hidden layers	1	2	2	3	3
Number of hidden nodes	200	300, 500	1200, 1200	400, 800, 1200	2000, 2000, 2000
Activation functions	ReLU	Linear, Linear	ReLU, ReLU	Tanh, Tanh, Sigmoid	ReLU, ReLU, ReLU

Accuracy evaluation

Instead of including all 400 accuracy measurements in this report, I'm going to present the results in a more compressed way with the help of statistical measures: min, max, median, average and variance. Also, for the sake of readability, I've marked the highest values of max, median and average in **bold**.

Training set

Network id	1	2	3	4	5
Min	0,97544	0,92384	0,99754	0,95898	0,99946
Max	0,97744	0,92668	0,99856	0,96136	**0,99994**
Avg	0,976749	0,925283	0,998247	0,960089	**0,999818**
Median	0,97683	0,92499	0,99828	0,95988	**0,99986**
Variance	3,44778E-07	5,54E-07	5,11E-08	4,87E-07	1,18E-08

Validation set

Network id	1	2	3	4	5
Min	0,974	0,9306	0,9859	0,961	0,9867
Max	0,9762	0,9317	0,9874	0,9623	**0,9882**
Avg	0,97502	0,93116	0,986645	0,961715	**0,98727**
Median	0,975	0,9312	0,9866	0,96185	**0,98725**
Variance	3,154E-07	1,08E-07	1,15E-07	1,4E-07	1,6E-07

Demo preparation set

Network id	1	2	3	4	5
Min	1	1	0,9	0,9667	0,8
Max	1	1	0,9666	1	0,9333
Avg	1	1	0,928305	0,998335	0,88331
Median	1	1	0,9333	1	0,9
Variance	0	0	0,000266	5,54E-05	0,001345

Network 1: 1 hidden layer, 200 ReLU

Although the number of nodes is not that high, and there's only one layer of hidden nodes, this network performs surprisingly well. The accuracy is above 97 % for all three sets, which means that it would probably get a very high number of points on the tests on the demo day.

Network 2: 2 hidden layers: 300 linear, 500 linear

This network configuration performed worse than network 1. I suspect that the reason is the linear activation function. It does not perform as well as the rectified linear unit activation function for this problem. It performs well on the demo set, though.

Network 3: 2 hidden layers: 1200 ReLU, 1200 ReLU

This network compares to network 1, but there's an important different: Because it has more nodes, it has specialized more on the training set. While performing well on the quite similar validation set, it does not perform so well on the demo set. I guess it pays too much attention to details, and that's bad because the demo preparation set has more noise (and thus more "details") than the training and validation set.

Network 4: 3 hidden layers: 400 tanh, 800 tanh, 1200 sigmoid

This network seems to be good at the demo preparation set. I guess that means that this network is more resistant to outliers. The images in the demo prep set also don't look much like the images in the training set. That must mean that this network is good at generalizing. Also, the accuracy is around 96 % on the training set and the validation set.

Network 5: 3 hidden layers: 2000 ReLU, 2000 ReLU, 2000 ReLU

This network configuration had the best performance on the validation set, but it didn't perform so well on the demo preparation set. Because of the relatively high number of layers and nodes, it has the capacity to get really well acquainted with the training set, and it starts to memorize each image rather than learn general trends. When it comes to a completely different set, its performance is mediocre. This is an example of overfitting. I will certainly not use this network configuraton on the demo day.