IT3105 Project module 5

Deep Learning for Image Classification

# 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.

In order to get consistent results, a pseudo random number generator (PRNG) is used for random values. These random values are used for things like initial weights and dropout. 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.

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.

I’ve used dropout in order to prevent overfitting. There’s 20 % input layer dropout and 50 % dropout on each hidden layer.

## One hidden layer, 200 nodes, rectified linear unit

Although the number of nodes is not that high, and there’s only one layer of hidden nodes, this network performs surprisingly well.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Seed | Training set accuracy | Validation set accuracy | Test set accuracy | Demo preparation set accuracy |
| 1 | 0.97676 | 0.9745 | 0.9734 | 1.0 |
| 2 | 0.97664 | 0.9762 | 0.9726 | 1.0 |
| 3 | 0.97684 | 0.9754 | 0.9736 | 1.0 |
| 4 | 0.9774 | 0.9746 | 0.9717 | 1.0 |
| 5 | 0.97544 | 0.9748 | 0.97 | 1.0 |
| 6 | 0.9771 | 0.9753 | 0.9733 | 1.0 |
| 7 | 0.97662 | 0.975 | 0.9728 | 1.0 |
| 8 | 0.97716 | 0.9741 | 0.9722 | 1.0 |
| 9 | 0.9773 | 0.9758 | 0.9739 | 1.0 |
| 10 | 0.97644 | 0.9751 | 0.9718 | 1.0 |
| 11 | 0.97744 | 0.9748 | 0.9732 | 1.0 |
| 12 | 0.97702 | 0.9756 | 0.9729 | 1.0 |
| 13 | 0.9764 | 0.9744 | 0.9715 | 1.0 |
| 14 | 0.97658 | 0.9748 | 0.9724 | 1.0 |
| 15 | 0.97544 | 0.974 | 0.9728 | 1.0 |
| 16 | 0.976 | 0.975 | 0.9714 | 1.0 |
| 17 | 0.97732 | 0.9748 | 0.9729 | 1.0 |
| 18 | 0.97734 | 0.9752 | 0.9722 | 1.0 |
| 19 | 0.97682 | 0.9754 | 0.9724 | 1.0 |
| 20 | 0.97692 | 0.9756 | 0.9723 | 1.0 |

## Two hidden layers with 300 and 500 nodes respectively, linear

This network configuration performed worse than the first one. 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Seed | Training set accuracy | Validation set accuracy | Test set accuracy | Demo preparation set accuracy |
| 1 | 0.92472 | 0.9313 | 0.9246 | 1.0 |
| 2 | 0.92474 | 0.9306 | 0.9251 | 1.0 |
| 3 | 0.92668 | 0.9309 | 0.923 | 1.0 |
| 4 | 0.9247 | 0.9312 | 0.9228 | 1.0 |
| 5 | 0.92586 | 0.931 | 0.925 | 1.0 |
| 6 | 0.92658 | 0.9315 | 0.9242 | 1.0 |
| 7 | 0.92496 | 0.931 | 0.9232 | 1.0 |
| 8 | 0.92608 | 0.9316 | 0.9237 | 1.0 |
| 9 | 0.92384 | 0.9306 | 0.9198 | 1.0 |
| 10 | 0.92622 | 0.931 | 0.9234 | 1.0 |
| 11 | 0.92484 | 0.9311 | 0.9234 | 1.0 |
| 12 | 0.92484 | 0.9312 | 0.9242 | 1.0 |
| 13 | 0.92574 | 0.9312 | 0.9254 | 1.0 |
| 14 | 0.92602 | 0.9314 | 0.9241 | 1.0 |
| 15 | 0.92512 | 0.9315 | 0.9245 | 1.0 |
| 16 | 0.92502 | 0.9317 | 0.9242 | 1.0 |
| 17 | 0.92488 | 0.9313 | 0.9236 | 1.0 |
| 18 | 0.92478 | 0.931 | 0.9228 | 1.0 |
| 19 | 0.92478 | 0.9306 | 0.9226 | 1.0 |
| 20 | 0.92526 | 0.9315 | 0.9242 | 1.0 |

## Two hidden layers with 1200 nodes each, rectified linear unit

This network configuration performed worse than the first one. 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Seed | Training set accuracy | Validation set accuracy | Test set accuracy | Demo preparation set accuracy |
| 1 | 0.99754 | 0.9865 | 0.9849 | 0.9333 |
| 2 | 0.99834 | 0.9865 | 0.9854 | 0.9333 |
| 3 | 0.99832 | 0.9865 | 0.9845 | 0.9333 |
| 4 | 0.99804 | 0.9863 | 0.9848 | 0.9 |
| 5 | 0.9983 | 0.9865 | 0.9852 | 0.9333 |
| 6 | 0.99816 | 0.9867 | 0.9851 | 0.9333 |
| 7 | 0.99848 | 0.9866 | 0.9856 | 0.9 |
| 8 | 0.99836 | 0.9866 | 0.9859 | 0.9333 |
| 9 | 0.9982 | 0.9868 | 0.9846 | 0.9333 |
| 10 | 0.99818 | 0.9865 | 0.9851 | 0.9333 |
| 11 | 0.99812 | 0.9871 | 0.9854 | 0.9 |
| 12 | 0.99856 | 0.9874 | 0.9858 | 0.9333 |
| 13 | 0.99822 | 0.9867 | 0.9852 | 0.9333 |
| 14 | 0.99842 | 0.9862 | 0.9858 | 0.9 |
| 15 | 0.99826 | 0.9868 | 0.985 | 0.9333 |
| 16 | 0.9982 | 0.9859 | 0.985 | 0.9333 |
| 17 | 0.9983 | 0.987 | 0.9853 | 0.9333 |
| 18 | 0.99798 | 0.9864 | 0.9863 | 0.9333 |
| 19 | 0.99848 | 0.987 | 0.9863 | 0.9666 |
| 20 | 0.99848 | 0.9869 | 0.9856 | 0.9333 |

## Three hidden layers with 400, 800, 1200 nodes and tanh, tanh, sigmoid respectively

This network seems to be particularly good at the demo preparation set, which is pretty noisy. 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Seed | Training set accuracy | Validation set accuracy | Test set accuracy | Demo preparation set accuracy |
| 1 | 0.95968 | 0.9614 | 0.9571 | 1 |
| 2 | 0.96096 | 0.9622 | 0.9582 | 1 |
| 3 | 0.96136 | 0.9623 | 0.9581 | 1 |
| 4 | 0.9599 | 0.9619 | 0.9564 | 1 |
| 5 | 0.96112 | 0.9614 | 0.9595 | 1 |
| 6 | 0.9595 | 0.9612 | 0.9575 | 1 |
| 7 | 0.95898 | 0.9619 | 0.9573 | 1 |
| 8 | 0.9603 | 0.962 | 0.9574 | 1 |
| 9 | 0.96104 | 0.9617 | 0.9595 | 1 |
| 10 | 0.95972 | 0.9619 | 0.9571 | 1 |
| 11 | 0.96044 | 0.9617 | 0.9577 | 1 |
| 12 | 0.9609 | 0.9621 | 0.9584 | 1 |
| 13 | 0.95964 | 0.9612 | 0.9582 | 1 |
| 14 | 0.95962 | 0.9619 | 0.9582 | 1 |
| 15 | 0.95956 | 0.962 | 0.957 | 1 |
| 16 | 0.96018 | 0.9618 | 0.9583 | 1 |
| 17 | 0.9592 | 0.9612 | 0.9562 | 0.9667 |
| 18 | 0.95986 | 0.9615 | 0.956 | 1 |
| 19 | 0.95938 | 0.961 | 0.9575 | 1 |
| 20 | 0.96044 | 0.962 | 0.9595 | 1 |

## Three hidden layers with 2000 nodes each, rectified linear unit

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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Seed | Training set accuracy | Validation set accuracy | Test set accuracy | Demo preparation set accuracy |
| 1 | 0.9998 | 0.987 | 0.9866 | 0.9333 |
| 2 | 0.99988 | 0.9867 | 0.9867 | 0.9 |
| 3 | 0.99994 | 0.9873 | 0.9861 | 0.9 |
| 4 | 0.9999 | 0.9873 | 0.9858 | 0.8666 |
| 5 | 0.99986 | 0.9877 | 0.9853 | 0.9 |
| 6 | 0.99946 | 0.9875 | 0.9853 | 0.9 |
| 7 | 0.99982 | 0.9869 | 0.9864 | 0.9333 |
| 8 | 0.99988 | 0.9872 | 0.9854 | 0.8333 |
| 9 | 0.99988 | 0.9877 | 0.9861 | 0.8666 |
| 10 | 0.99988 | 0.9878 | 0.9855 | 0.8333 |
| 11 | 0.9997 | 0.9868 | 0.9858 | 0.8 |
| 12 | 0.9998 | 0.9868 | 0.9853 | 0.9 |
| 13 | 0.99986 | 0.9875 | 0.9855 | 0.9 |
| 14 | 0.99986 | 0.9871 | 0.9868 | 0.8666 |
| 15 | 0.99984 | 0.987 | 0.9848 | 0.9 |
| 16 | 0.99988 | 0.9876 | 0.9864 | 0.9 |
| 17 | 0.99984 | 0.9882 | 0.9864 | 0.9333 |
| 18 | 0.99988 | 0.9873 | 0.9862 | 0.8666 |
| 19 | 0.99974 | 0.9872 | 0.9865 | 0.8333 |
| 20 | 0.99966 | 0.9868 | 0.9858 | 0.9 |