Hackathon1 - Creating and Training Models

Beichen Zhang

In this assignment, to investigate the impacts of different architectures of the neural network and some hyperparameters, such as the number of neurons, the number of layers, different activation functions, and different optimizers. I used the a function in the TensorFlow: tf.keras.layers to control the number of layers and the number of neurons in each layer (Table 1). Then, I created different models on the training data set which is 90% of the MNIST data set. I set the maximum epoch as 20 with an early-stop parameter of 0.01, which means the training will stop when the difference of accuracy on the training data set between two epochs is smaller than 0.01. The loss function was calculated by the difference between the logits and labels using the sparse softmax cross entropy. A gridded research was used to find the best structure. Single-hidden-layer and two-hidden-layer structures were tested in this assignment, along with a list of the number of neurons: 50, 100, 200. The activation function was same for every hidden layer. And ReLu, Sigmoid, and Tanh were tested. The training time of every architecture was also outputted.

Hyperparameter	Values
# hidden layers	[1,2]
# neurons	[200,100,50]
# epoch and early-stop	20, 0.01
Activation function	[relu, sigmoid, tanh]

Table 1: Gridded searching options for training the architecture of the neural network.

#neurons of #1 layer	#neurons of #2 layer	# stopped epoch	Activation function	Training time (Seconds)	Validation Accuracy
200	200	4	Relu	34.22	0.9500
200	100	4	Relu	33.08	0.9582
200	50	6	Relu	48.12	0.9555
100	200	6	Relu	47.78	0.9522
100	100	4	Relu	33.48	0.9505
100	50	5	Relu	39.53	0.9462
50	200	5	Relu	39.68	0.9138
50	100	5	Relu	39.26	0.9182
50	50	5	Relu	38.15	0.9430

Table 2: Gridded search for the two-hidden-layer neural network with ReLu activation function

Training results of the two-layer neural network with the ReLu activation function are presented in Table 2. The optimum two-hidden-layer structure with 200 neurons in the first hidden layer and 100 neurons in the second hidden layer had the best validation accuracy (0.9582). Training results of the two-layer neural network with the Sigmoid activation function are presented in Table 3. The optimum two-hidden-layer structure with 200 neurons in the first hidden layer and 100 neurons in the second hidden layer had the best validation accuracy (0.9137). Training results of the two-layer neural network with the Tanh activation function are presented in Table 4. The optimum two-hidden-layer structure with 200 neurons in the first hidden layer and 100 neurons in the second hidden layer had the best validation accuracy (0.8973). Above all, the two-hidden-layer network had the same number of neurons for its first and second hidden layer as the optimum model. And the network with the ReLu activation function was the best performed model that had the highest validation accuracy (0.9582).

#neurons of #1 layer	#neurons of #2 layer	# stopped epoch	Activation function	Training time (Seconds)	Validation Accuracy
200	200	4	Sigmoid	34.32	0.9208
200	100	4	Sigmoid	33.54	0.9137
200	50	4	Sigmoid	33.00	0.9128
100	200	3	Sigmoid	24.48	0.8813
100	100	4	Sigmoid	32.89	0.8907
100	50	5	Sigmoid	40.03	0.9017
50	200	3	Sigmoid	23.26	0.8797
50	100	5	Sigmoid	38.68	0.8962
50	50	4	Sigmoid	29.95	0.8907

Table 3: Gridded search for the two-hidden-layer neural network with Sigmoid activation function

#neurons of #1 layer	#neurons of #2 layer	# stopped epoch	Activation function	Training time (Seconds)	Validation Accuracy
200	200	4	Tanh	33.14	0.8938
200	100	3	Tanh	24.98	0.8973
200	50	4	Tanh	32.53	0.8945
100	200	4	Tanh	33.08	0.8792
100	100	5	Tanh	41.04	0.8847
100	50	4	Tanh	32.15	0.8723
50	200	5	Tanh	39.16	0.8815
50	100	5	Tanh	39.25	0.8687
50	50	6	Tanh	45.86	0.8592

Table 4: Gridded search for the two-hidden-layer neural network with Tanh activation function

The training process of the single-hidden-layer neural network was similar to the two-hidden-layer one. The training time of the single-hidden-layer neural network was significantly lower than the two-hidden-layer. However, the validation accuracy also decreased. The best performed model was still the network with the ReLu activation function. And the more neurons we applied to the hidden layer, the better performance of the neural network. So the 200 neurons network has the highest validation accuracy (0.9400).

#neurons of #1 layer	# stopped epoch	Activation function	Training time (Seconds)	Validation Accuracy
200	4	ReLu	29.84	0.9400
100	5	ReLu	33.18	0.9322
50	6	ReLu	38.73	0.9080
200	3	Sigmoid	20.60	0.9127
100	3	Sigmoid	20.22	0.8888
50	4	Sigmoid	26.76	0.8732
200	4	Tanh	27.81	0.8908
100	6	Tanh	40.75	0.8890
50	6	Tanh	38.81	0.8478

Table 5: Gridded search for the single-hidden-layer neural network with ReLu, Sigmoid, and Tanh activation function