Cairo University

Faculty of Computers and Artificial Intelligence

AI322: Supervised Learning

Assignment 3

Name	ID	Phone Number
Mahmoud Hossam Atef	20180251	01000657682
Abdelrahman Khaled Ahmed Ali	20180142	01099342630

Delivery: 18-05-2021

Introduction

The base configurations for our model are 2 convolution layers (32 and 64 filters 2x2) and 2 fully connected layers (64 and 10) with total 13,226 parameters. Using RelU activation function, SGD as the optimizer, 32 batch size, and 10 epochs.

We used a Google sheet to save the required data in, it is available at the end of the report and you can also view it here.

Choosing Epochs

We choose epochs as 20 because it has the highest test accuracy compared to 10, 13, and 15 epochs.

Choosing Learning Rate

We choose 0.05 as learning rate as it has the highest training accuracy while maintaining a small difference between training and testing accuracy (avoiding overfitting), and while reaching minimum value in SGD fast.

Editing Number of Filters and Neurons in Layers

We tried various combinations of number of filters in the convolution layers and number of neurons in the fully connected layers. We found that the model that had 3 convolution layers all consisting of 32 filters and 2 fully connected layers with 32 and 10 neurons respectively had the highest accuracy (97.77%, row 17) and it had 9,802 parameters. On the other hand, the model that had 3 convolution layers all consisting of 16 filters and 2 fully connected layers with 16 and 10 neurons respectively had a 2% lower accuracy (95.46%, row 18) but it had the lowest parameters 2,602.

We neglected the parameters and we had chosen to stick with the model having the highest accuracy; all next trials were made upon the model having the highest accuracy. Later, we were able to adjust the second model that made it achieve 97% accuracy with much less parameters than the first model (2,682 only, row 44).

Choosing Batch Size

We tried both 64 and 96 batch, and we found that 96 batch size is better as it decreased the average training time from 11 seconds to 8 seconds while maintaining the same accuracy. We will stick with this model (row 24) using 96 batch size and we will continue our trials on it.

Choosing Activation Function

We tried three other activation functions other that ReLU, the Softplus, SoftSign, and Sigmoid. But all of them lowered the accuracy our model accuracy. Thus, we will just keep ReLU as the activation function.

The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. Also, ReLU helps to prevent the exponential growth in the computation required to operate the neural network. If the CNN scales in size, the computational cost of adding extra ReLUs increases linearly.

We also noticed that the sigmoid activation function doesn’t work well with the optimizer being SGD because sigmoid outputs are not zero-centered, which is undesirable because it can indirectly introduce undesirable zig-zagging dynamics in the gradient updates for the weights.

Choosing the Optimizer

We tried two different optimizers other than SGD, the Adadelta and Adam. Both optimizers reduced our model’s accuracy significantly and we decided to keep SGD as our optimizer.

Also, SGD is easier to fit in the memory due to a single training example being processed by the network. It is computationally fast as only one sample is processed at a time. For larger datasets, it can converge faster as it causes updates to the parameters more frequently.

Due to frequent updates, the steps taken towards the minima of the loss function have oscillations that can help to get out of the local minimums of the loss function (in case the computed position turns out to be the local minimum).

Dropout layers

We tried implementing dropout layer in different positions within the network with different dropout rates but it just reduced both training and test accuracy. We decided not to add a dropout layer for our model.

Conclusion

Having our model with best number of convolution layers, number of fully connected layers, batch size, activation function, and optimizer, we decided to change the filters’ kernel in our model trying to achieve better accuracy (rows from 43 to 50). The changes we made weren’t satisfying; an increase in test accuracy by 1% required increase in parameters by 56% which use too much resources.

We then tried to improve the model that had the lowest parameters, the model had 2,602 parameters (row 18 but with 96 batch size) and we decided to edit its filter kernels (rows from 54 to 60). We were able to improve its test accuracy by 2% (95% to 97%) that required 3% more parameters (80 more parameters). This model is definitely better than the previous model as it achieved the same 97% test accuracy with 72% less parameters. We decided to go with this model and the provided code is for it as well.

CNN Layers	Filters	FC Layers	FC Layers Neurons	Parameters	Activation Used	Learning Rate	Dropout layer Position	Dropout Rate	Optimizer	Epochs	Batch Size	Training Avg Time	First 5 Epochs Accuracy	Training Accuracy	Testing Avg Time	Test Accuracy	Train & Test Accuracy Difference
2	32,64	2	64,10	13,226	RelU	0.01	0	0	SGD	10	32	11s	0.227, 0.5659, 0.6292, 0.6882, 0.7327	0.8298	1s	0.8447	0.0149
2	32,64	2	64,10	13,226	RelU	0.01	0	0	SGD	13	32	11s	0.2443, 0.5271, 0.6139, 0.6671, 0.7142	0.8298	1s	0.8447	0.0149
2	32,64	2	64,10	13,226	RelU	0.01	0	0	SGD	15	32	11s	0.3379, 0.5901,0.6727, 0.7192, 0.7142	0.8442	1s	0.8548	0.0106
2	32,64	2	64,10	13,226	RelU	0.01	0	0	SGD	20	32	10s	0.2544, 0.5681, 0.6463, 0.6980, 0.7430	0.8515	1s	0.8647	0.0132
																	0
2	32,64	2	64,10	13,226	RelU	0.025	0	0	SGD	20	32	10s	0.4221, 0.6848, 0.7751, 0.8031, 0.8184	0.8759	1s	0.8786	0.0027
2	32,64	2	64,10	13,226	RelU	0.04	0	0	SGD	20	32	10s	0.5287, 0.7863, 0.8229, 0.8388, 0.8472	0.8852	1s	0.8755	-0.0097
2	32,64	2	64,10	13,226	RelU	0.05	0	0	SGD	20	32	10.7s	0.4620, 0.7550,0.8045,0.8245, 0.8371	0.8837	1s	0.8764	-0.0073
2	32,64	2	64,10	13,226	RelU	0.1	0	0	SGD	20	32	10s	0.3387, 0.7203, 0.8200, 0.8228, 0.8471	0.8892	1s	0.8737	-0.0155
																	0
2	32,32	2	64,10	7,050	RelU	0.05	0	0	SGD	20	32	10s	0.2104, 0.5099, 0.5735, 0.6146, 0.6517	0.8117	1s	0.8246	0.0129
2	64,32	2	64,10	7,050	RelU	0.05	0	0	SGD	20	32	14s	0.4330,0.7395, 0.7950, 0.8131, 0.8267	0.8708	1s	0.8606	-0.0102
2	32,64	2	32,10	10,826	RelU	0.05	0	0	SGD	20	32	11s	0.4691, 0.7530, 0.8091, 0.8237, 0.8353	0.8774	1s	0.8682	-0.0092
2	32,32	2	32,10	5,674	RelU	0.05	0	0	SGD	20	32	9.5s	0.4289, 0.7354, 0.7907, 0.8092, 0.8233	0.8643	1s	0.8632	-0.0011
																	0
3	32,32,32	2	32,10	9,802	RelU	0.05	0	0	SGD	20	32	11s	0.5771, 0.9129, 0.9383, 0.9509, 0.9572	0.9812	1s	0.9777	-0.0035
3	16,16,16	2	16,10	2,602	RelU	0.05	0	0	SGD	20	32	8.2s	0.1533, 0.7021, 0.7960, 0.8413, 0.8746	0.9519	1s	0.9546	0.0027
3	32,32,16	2	16,10	6,794	RelU	0.05	0	0	SGD	20	32	11s	0.5928, 0.9056, 0.9356, 0.9439, 0.9507	0.9749	1s	0.9618	-0.0131
3	32,32,32	3	16,16,10	9,386	RelU	0.05	0	0	SGD	20	32	11s	0.5653, 0.9053, 0.9367, 0.9513, 0.9584	0.9801	1s	0.9715	-0.0086
3	32,32,32	3	32,16,10	10,170	RelU	0.05	0	0	SGD	20	32	11s	0.5120, 0.9040, 0.9376, 0.9499, 0.9584	0.9819	1s	0.9746	-0.0073
3	32,32,32	2	32,10	9,802	RelU	0.05	0	0	SGD	20	64	9.05s	0.44889, 0.8642, 0.9082, 0.9292, 0.9394	0.9742	1s	0.9703	-0.0039
3	32,32,32	2	32,10	9,802	RelU	0.05	0	0	SGD	20	96	8.05s	0.4094, 0.8412, 0.8902, 0.9108, 0.9248	0.9709	1s	0.97	-0.0009
3	32,32,32	2	32,10	9,802	Softplus	0.05	0	0	SGD	20	96	9.15s	0.1034, 0.1248,0.5774, 0.8073, 0.8520	0.9503	1s	0.9502	-0.0001
3	32,32,32	2	32,10	9,802	SoftSign	0.05	0	0	SGD	20	96	8s	0.3479, 0.7591, 0.8575, 0.8982, 0.9179	0.9662	1s	0.9634	-0.0028
3	32,32,32	2	32,10	9,802	Sigmoid	0.05	0	0	SGD	20	96	8s	0.1126, 0.1072, 0.1114, 0.1085, 0.1063	0.1088	1s	0.1135	0.0047
3	32,32,32	2	32,10	9,802	RelU	0.05	0	0	Adadelta	20	96	8.2s	0.1317, 0.6576, 0.7612, 0.7839, 0.8079	0.9049	1s	0.9092	0.0043
3	32,32,32	2	32,10	9,802	RelU	0.05	0	0	Adam	20	96	6.75	0.6472, 0.8107, 0.8212, 0.8287, 0.8316	0.8458	1s	0.8406	-0.0052
3	32,32,32	2	32,10	9,802	RelU	0.05	3	0.2	SGD	20	96	8.05s	0.3230, 0.7716, 0.8336, 0.8699, 0.8885	0.9446	1s	0.9646	0.02
3	32,32,32	2	32,10	9,802	RelU	0.05	3	0.4	SGD	20	96	8.25s	0.3060, 0.7130, 0.8017, 0.8397, 0.8598	0.9173	1s	0.9606	0.0433
3	32,32,32	2	32,10	9,802	RelU	0.05	4	0.2	SGD	20	96	8.05s	0.4179, 0.7782, 0.8377,0.8700, 0.8904	0.9524	1s	0.9672	0.0148
3	32,32,32	2	32,10	9,802	RelU	0.05	4	0.3	SGD	20	96	8.05s	0.3052, 0.7382, 0.8028, 0.8440, 0.8733	0.9389	1s	0.9649	0.026
3	32,32,32	2	32,10	9,802	RelU	0.05	4	0.4	SGD	20	96	8.05s	0.3064,0.6988, 0.7750, 0.8154, 0.8484	0.9273	1s	0.96	0.0327

Trying different filter sizes and CNN layers in the model in Row 24 and maintaining the same activation function, optimizer, learning rate, dropout rates, batch size,and number of epochs.

CNN Layers	Filters	FC Layers	FC Layers Neurons	Parameters	Training Avg Time	Training Accuracy	Testing Avg Time	Test Accuracy	Train & Test Accuracy Difference	Increase in parameters	Increase In Test accuraccy
3	32(2x2), 32(2x2), 32(2x2)	2	32,10	9,802	8.05s	0.9709	1s	0.97	-0.0009	~ Original Model ~	~ Original Model ~
3	32(4x4), 32(3x3), 32(2x2)	2	32,10	15,306	11s	0.9904	1s	0.981	-0.0094	56.15%	1.13%
3	32(3x3), 32(3x3), 32(2x2)	2	32,10	15,082	9s	0.9875	1s	0.9833	-0.0042	53.86%	1.37%
3	32(3x3), 32(3x3), 32(2x2)	2	16,10	14,394	9s	0.9884	1s	0.9834	-0.005	46.84%	1.38%
3	32(3x3), 32(2x2), 32(2x2)	2	32,10	9,962	8.4s	0.9845	1s	0.9792	-0.0053	1.63%	0.94%

Trying different filter sizes and CNN layers in the model in Row 18 with batch size = 96 and maintaining the same activation function, optimizer, learning rate, dropout rates,and number of epochs.

CNN Layers	Filters	FC Layers	FC Layers Neurons	Parameters	Training Avg Time	Training Accuracy	Testing Avg Time	Test Accuracy	Train & Test Accuracy Difference	Increase in parameters	Increase In Test accuraccy
3	16(2x2), 16(2x2), 16(2x2)	2	16,10	2,602	8.2s	0.9519	1s	0.9546	0.0027	~ Original Model ~	~ Original Model ~
3	16(3x3), 16(2x2), 16(2x2)	2	16,10	2,682	6.05s	0.9717	1s	0.97	-0.0017	3.07%	1.61%
3	16(3x3), 16(3x3), 16(2x2)	2	16,10	3,962	6.9s	0.9795	1s	0.9746	-0.0049	52.26%	2.06%
3	16(4x4), 16(2X2), 16(3X3)	2	16,10	4,074	7.25S	0.9783	1S	0.9782	-0.0001	56.57%	2.47%
									0		