

Extended MNIST Dataset - Convolutional Network

Homework 04

Kalin Gibbons

November 10, 2020

Model Summary

The final model achieved an accuracy of 88.2% on the test set, with a best validation accuracy of 88.5%. It consisted of 5 layers with a total of 671,791 trainable parameters (Table 1). The first two layers were 2D convolutional layers with max-pooling, followed by a third 2D convolutional layer and 10% dropout. The output of this last convolutional layer was flattened and put through a densely-connected layer. Finally, a 50% dropout was applied before the final softmax classification layer.

The model was trained using the ADAM optimizer with a categorical crossentropy loss function, using a learning rate of 5×10^{-4} . Data augmentation was performed using a randomized zoom of $\pm 20\%$, and a rotation of $\pm 5^\circ$. An identically augmented set representing 10% of the shuffled EMNIST dataset was withheld for validation, and the model was trained using a mini-batchsize of 8 images, for 15 epochs. An early-stopping criteria was implemented to stop training after 5 epochs with no improvement in validation accuracy, but was never triggered.

Table 1. Network architecture for EMNIST convolutional neural network.

Layer	Output Shape	Main Attribute	Kernel-/Pool- Size	Activation	Neurons	Param #
Input	8x24x24	-	-	-	4608	0
Conv2D	24x24x64	64	5x5	ReLU	36,864	1,664
Pool2D	12x12x64	Max	2x2	-	9,216	0
Conv2D	10x10x128	128	3x3	ReLU	12,800	73,856
Pool2D	5x5x128	Max	2x2	-	3,200	0
Conv2D	3x3x256	256	2x2	ReLU	2,304	295,168
Dropout	3x3x256	0.1	-	-	2,304	0
Flatten	2304	Ravel	-	-	2,304	0
Dense	128	128	-	ReLU	128	295040
Dropout	128	0.5	-	-	128	0
Dense	47	47	-	Softmax	47	6063

Conclusions

The EMNIST dataset was already centered and normalized, so implementing any shift augmentations were not applicable to this dataset, but they should be useful for other image recognition tasks. Likewise, flipping the data along either axis would hinder the networks ability to correctly classify characters. A slight zoom

(20%) and rotation (5°) was selected to artificially enlarge the dataset. This was a reasonable selection because people are likely to write in italics with a variety of slants, and the zoom changes the active pixels while maintaining the same character aspect ratio. Most examples of Convolutional Neural Networks for classification tasks have an increase in depth, followed by a contraction in depth, and this network followed suit. The numbers of filters were selected to be powers of two, as the training was performed on a GPU, which are optimized for base-2 memory access.