Submission for Deep Learning Exercise

Team: shallow_learning_group
Students: Batuhan Karaca

January 9, 2024

Coding Tasks

1) Base Network

The accuracy was 32% as expected.

2) Increasing Number Of Filters

The accuracy increased to 37%. More parameters increased the networks flexibility. However, the increased flexibility can increase variance, resulting in overfitting. In our case, the number of parameters, when number of filters was 16, were so that the network was fit *less* to the training data compared to when number of filters was 3, thus performing better.

3) Reducing Filter Size

The accuracy remained the same. There may not be a correlation between the generalization accuracy and filter size. Let K, N_o, C_{in} and C_{out} be filter size, output size, number of channels in the input layer and number of channels in the output layer respectively. Then time complexity of convolution and space complexity of the filters in a layer become $O(K^2N^2C_{out})$ and $O(K^2C_{in}C_{out})$ respectively, which means that lower filter sizes are more time and memory efficient. Furthermore, larger filter sizes imply larger receptive field.

4) Inserting Batch Normalization

The accuracy dramatically increased to 48%. Regularization techniques bring bias to the model. Some of them achieve this by simplifying the model. As discussed in the Deep Learning book, batch normalization simplifies the model's output, making it to be more easily estimated as a standard Gaussian [1].

Model	Test Accuracy (%)
`ModifiedResNet1`	60
`ModifiedResNet2`	37
`ModifiedResNet3`	44
`ModifiedResNet4`	42
`ModifiedResNet5`	51

Table 1: Test accuracies of the models after the experiment

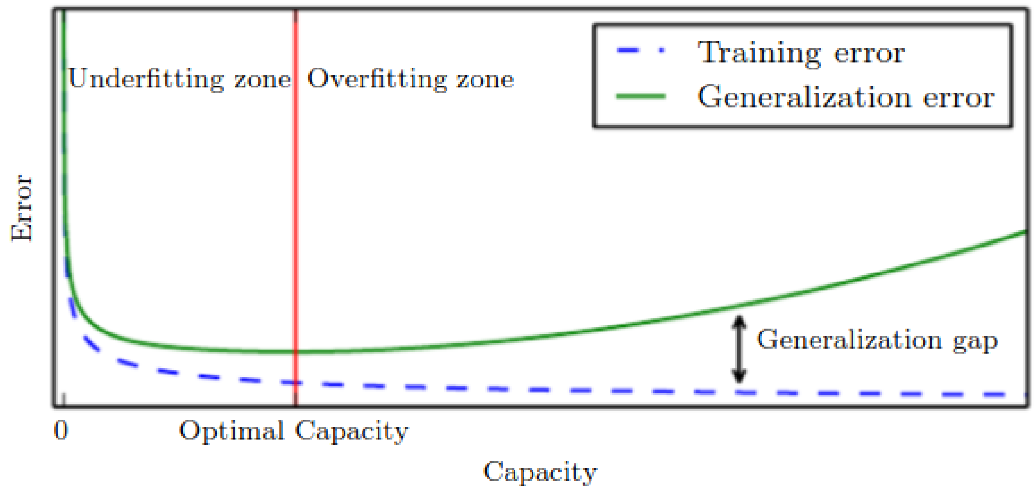

Figure 1: The error-capacity graph from the Deep Learning Book of Goodfellow et al (2016) [1]

5) Transfer Learning With Resnets

The results are given in table 1. The trend is that with more layers trained, the network performed mostly better. For the same data and same number of epochs, as more layers are fine tuned, the capacity of the network increases. The error-capacity graph is given in figure 1. The accuracy-capacity relation will be the upside-down version of this. In our case, the network does not seem to overfit yet, hence more fine-tuning should increase generalization accuracy. However, the generalization error is also influenced by the noise in the data. The exception in the transition from the third to the fourth network could be attributed to this.

6) ResNets handling variety sized inputs

Given number of parameters, convolution operation can have variety size of inputs compared to matrix multiplication. For given K, P and s as filter size, padding and stride respectively; the input size N_i can be any number such that the output size $N_o = \left\lfloor \frac{N_i+2P-K}{s} \right\rfloor + 1$ is greater than zero. In other words, $N_i \geq K-2P$, implying that input size can be arbitrarily large. On the contrary (assuming 2d weight matrices), given a number of parameters θ in a fully-connected layer, N_i can only be $\frac{\theta}{N_o}$. If the deeper layers of the networks are properly set up, a convolutional neural network (CNN), which starts with a convolutional layer can have inputs with arbitrary sizes compared to a feed-forward neural network (FFN). ResNet is a CNN model, hence is able to do so as well.

References

[1] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.