

Homework 7 ConvMixer network

Mihalache Radu-Stefan

November 25, 2023

1 Model structure

I chose to use a ConvMixer neural network because of its good results and the small number of trainable parameters. The ConvMixer makes use of patch embedding, batch normalisation, skip connections (residuals) and regular 2d convolutional layers.

The structure of the model is the following:

1. Conv2d; kernel_size=2; stride=2; This takes care of the patch encoding,
2. GELU (variation of RELU)
3. Batch Normalization (normalizes the distribution)
4. A list of the following layers
 1. Conv2d layer
 2. GELU
 3. Batch Normalization
8. A skip connection between 1 and 3
9. 4. Conv2d layer
10. 5. GELU
11. 6. Batch Normalization
12. After a list of the above, repeated a certain depth
13. AdaptiveAvgPool2d (An adaptation of average pooling that tries to automatically find the stride and kernel size)
14. A fully connected layer

2 Forward pass

The flow of the input signal goes from one Conv layer to a GELU activation function, batch normalization and some skip connections. The exact flow follows the list above. A Conv2d layer applies a filter over the matrix. Batch normalization normalizes the distribution of values. Average pooling does the average for each section of a matrix given a kernel size. AdaptiveAvgPool2d just automatically finds the best kernel size and stride. Finally, the data is fed through a fully connected neural network.

3 Backward pass

Gradients are computed for the fully connected layer and then propagated backwards towards the convolutional layers. Each convolutional layer computes the gradients just for the neurons in the applied filter. Given that ConvMixer makes use of skip connections, gradients are not lost given that the neural network is deep.

4 Layer contributions

The biggest contributions to the results come from the first conv layer that does patch encoding, the layers with skip connections that make it possible to efficiently train a deep neural network and the average pooling that reduces the dimensions of the data to be feed to the fully connected layer.