PA1

20120888 조민기

Model specification

Optimizer: Adam Optimizer

Loss: sparse_softmax_cross_entropy

Layer	Input	1	2	3	4	5	6	7
# of node	28x28	512	256	128	64	32	16	10

# of param

Depth 3: 5.35×10^5

Depth 5: 5.74×10^5

Depth 7: 5.77×10^5

Experiment

Depth 3: No dropout : 42.6%

start with 512 30.8%

Depth 5: with no dropout 39.85

Starting with 512 parameters / Dropout: 26%

Starting with 1024 parameteres / Dropout : 22%

Depth 7: 36.8%

	No dropout	With dropout (rate 0.5)	With dropout (rate 0.2)
Depth 3	44.4%	30.8%	33.8%
Depth 5	39.85%	26.3%	37.1%
Depth 7	36.8%	22.3%	30.9%

CNN with 2 conv layer, 1 FC layer(1024 feature) : 66%

Analize

- Accuracy goes worse when model gets deeper.
  AlexNet has 8 layer and VGG 16 or deeper CNN networks works better for image classification. According to these studies, we can say that deeper network can solve more complex problem. But in my experiment, this isn't the case. That's because FC layer is not good for image classification problem.
  Of course FC network also can solve same or even more difficult problem. But FC has too much information and because of curse of dimensionality, it gets hard to train the FC network. Recent studies say that FC network in CNN easily cause the overfitting, and that's why many

people are ridding off the FC layer from the network for image classification problem.
Double number of neuron also works worse. So we can say that too much parameter with too many syntactical levels makes model hard to get optimized.

- Model with dropout work worse even when it has a same number of parameter. And too much high rate also makes model bad.
  For Depth 5 network, model with no dropout has 39.85% of accuracy and model with dropout has 26% of accuracy. (dropout rate was 50%)
  Dropout for image classification could be bad because it makes model lose some local information from the image data. It's because Image data usually has its important feature in specific local area.
  Actually AlexNet or VGG net has dropout for the FC model. But that is fine because those layers deal with features which is already extracted by convolutional layers, so it does not lose local information for image data.

- Adam Optimizer converge faster than GradientDescent Optimizer.
  When 50step after start, Gradient descent algorithm goes 9.8% accuracy and Adam goes 10.5% accuracy