Hyomin Seo

Title : ``Rethinking Generalization'' Analyzation. Mobile, Squeeze and DenseNet

1.Describe one of the experiments in ``Rethinking Generalization'' and what the implications were.

Main experimentation

Some standard architectures where trained on a copy of data where the true labels were explained by random labels.

Results

- Deep neural networks can easily fit random labels.
- Deep neural networks generalize poses a stronger model capacity, and a feed-foward networks with a single hidden layer that contains a finite number of neurons can approximate continuous functions on compact subset.
- Stochastic Gradient Descent may perform an implicit regularizer.
- " for linear models, SGD always converse to a solution with the small norm, showing that the algorithm itself is implicitly regularizing the solution".

2. Implication ~ more like a conclusion statement based on observation made from experimentation

The results of the experimentation rules out some of the classical learning theories such as VC dimension, Rademacher complexity and uniform stability as possible explanations for the generalization performance of neural networks.

3. Definition

VC dimension

Vapnik–Chervonenkis theory VC dimension is a measure of the capacity (complexity, expressive power, richness, or flexibility) of a set of functions that can be learned by a statistical binary classification algorithm.

Rademacher complexity

measures the ability of an hypothesis class H to fit random ±1 binary labels. If compared to the VC dimension, Rademacher complexity is distribution dependent and defined for any class of real-valued functions (not only discrete-valued functions).

Uniform Stability

strong condition which is not met by all algorithms but is, surprisingly, met by the large and important class of Regularization algorithms. The generalization bound is given in the article.

4. Compare and contrast Squeezenet with MobileNets.

	SqueezeNet	**MobileNets**
Definition	A type of a Convolutional Neural Networks with a smaller size that offers various advantages for being 'Squeezed"	Model for mobile and embedded vision application that is structured on a streamlined architecture that use depth wise separable convolutions to build light weight deep neural networks. Architecture :
Architecture	Employes the two layers Squeeze layer : reduce the depth to a smaller number Expand layer : increases the depth. The squeeze layer and expand layer keep the same feature map size.,	Employes two convolutions (factorize a standard convolution into two separate convolutions) Depthwise convolution : applies a single filter to each input channel. 1X1 pointwise convolution : applies 1X1 convolution to combine the output the depthwise convolution.
Advantages	Less communication across servers during distributed training. Less bandwidth to export a new model from the cloud to an autonomous car. More feasible to deploy on FP-GAs and other hardware with limited memory.	It can generally reduce network size and number of parameters. It can work with small and low latency convolutional neural networks.

5. DenseNets?

DenseNets

A DenseNet is a type of convolutional neural network that utilises dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other. To preserve the feed-forward nature, each layer obtains additional inputs from all preceding layers and passes on its own feature-maps to all subsequent layers.

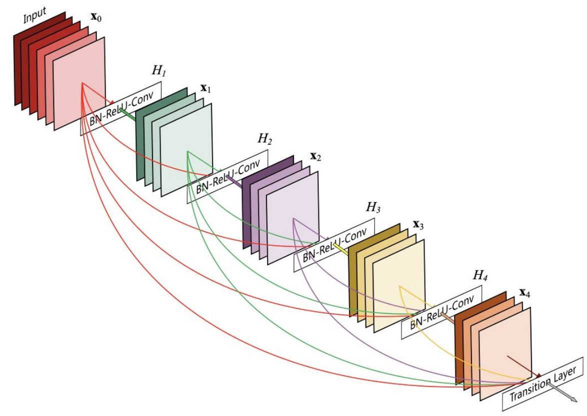

Figure 1: A 5-layer dense block with a growth rate of $k = 4$. Each layer takes all preceding feature-maps as input.

Dense net adds shortcuts among layers to to improve the information flow along layers; each layer in densenet receives all of the outputs of all of the previous layers, concatenate the received outputs in the depth dimension. Densenet stahcs the output layer over the depth dimension, giving efficient in terms of parameters (fewer parameters) and computation (same lever of accuracy compared to other model). However, from the paper *'SparseNet: A Sparse DenseNet for Image Classification'* it states that the exact concept that makes the DenseNet appealing is simultaneously the source of network prone to overfitting. This particular paper suggest to spartify the DenseNet; where some of the connection between the middle layers of Dnesenet is dropped to reduce connections of individual layers.

'SparseNet A Sparse DenseNet for Image Classification' *:*

https://arxiv.org/pdf/1804.05340.pdf