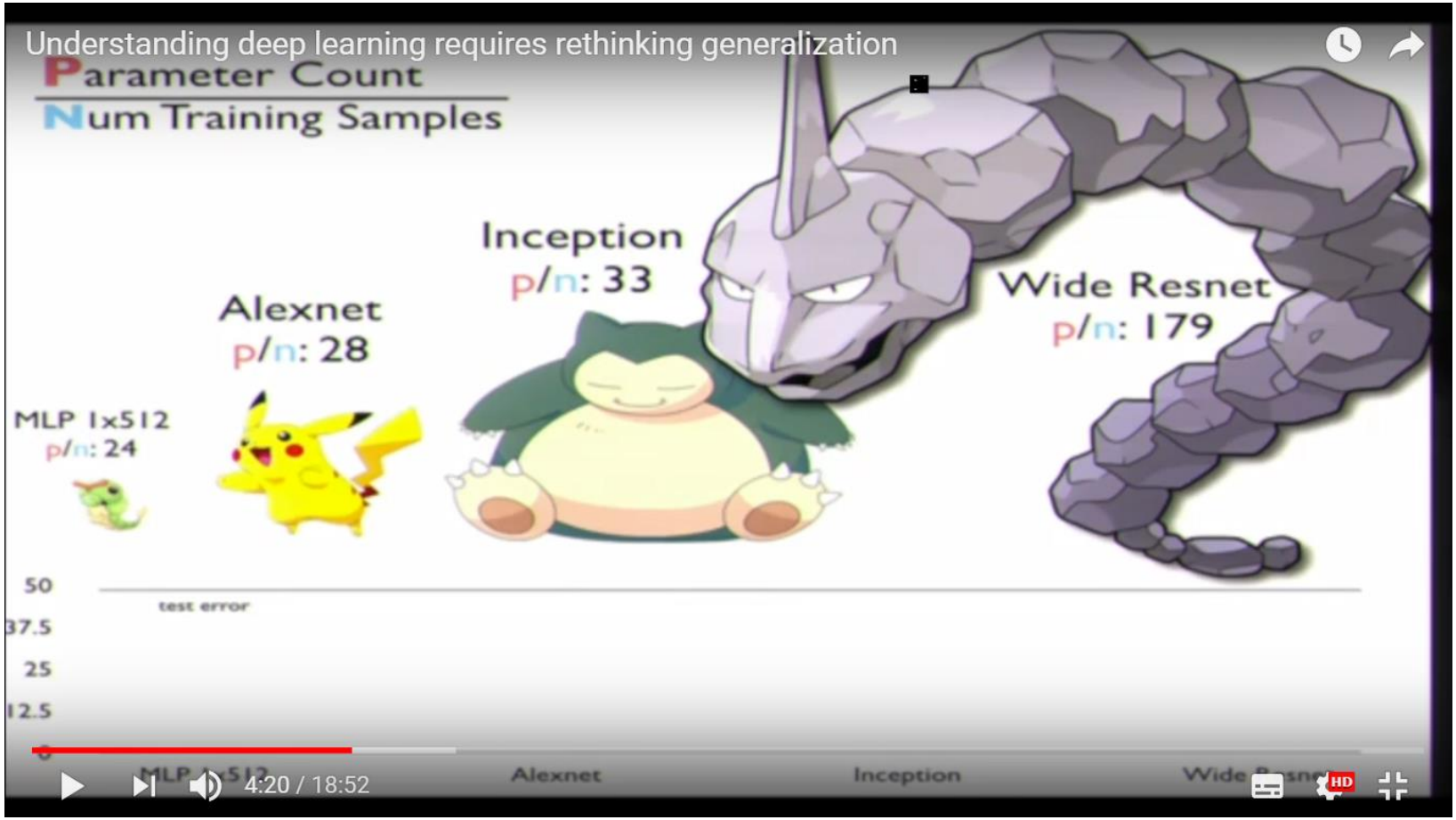
Generalization Ability

We use very large network today



Source of image: https://www.youtube.com/watch?v=kCj51pTQPKI

機器學習基石
https://youtu.be/XxPB9GlJEUk

Generalization Gap

No matter the data distribution

With probability $1-\delta$

$$E_{train} \leq E_{test} \leq E_{train} + \Omega(R, M, \delta)$$

Smaller δ, larger Ω

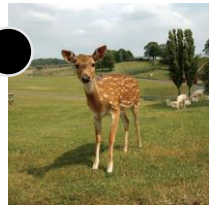
R is the number of training data → Larger R, smaller Ω

M is the "capacity" of your model → Larger M, larger Ω

("size" of the function set)

How to measure the "capacity"? VC dimension (d_{VC})

Given 3 data points



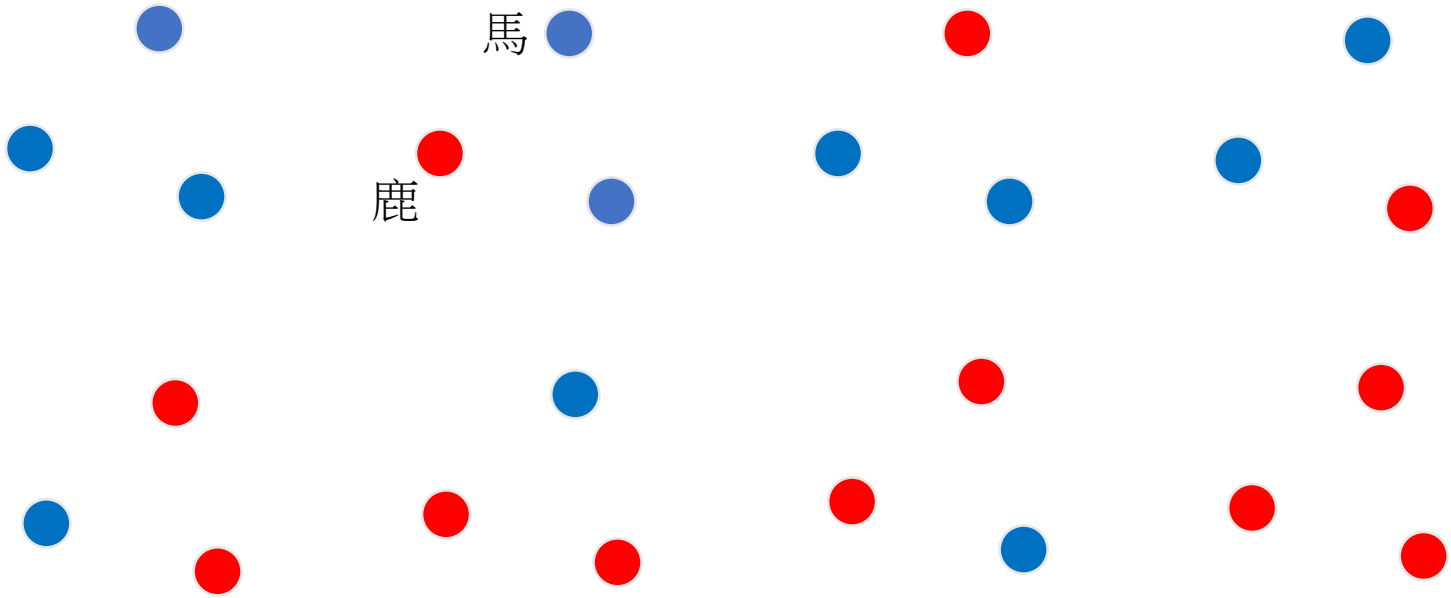
Random label (故意亂教)

Model M can always achieve 0% error rate

(亂教 Model M 都學得會)

VC dimension (d_{VC}) of Model M ≥ 3

e.g. linear model

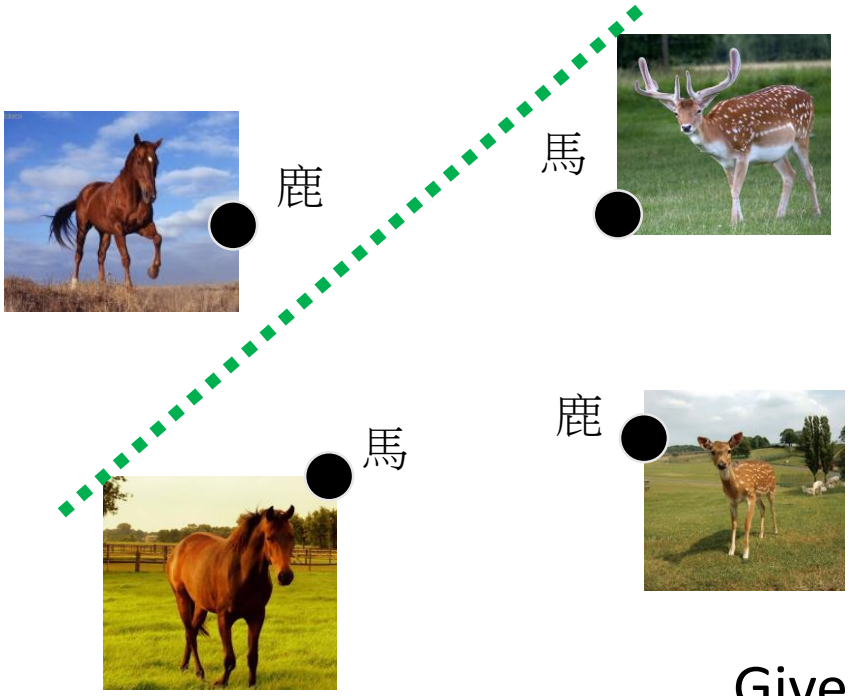


Random label (故意亂教)

There are some cases linear model can not learn.

(知道是來亂的，所以不學)

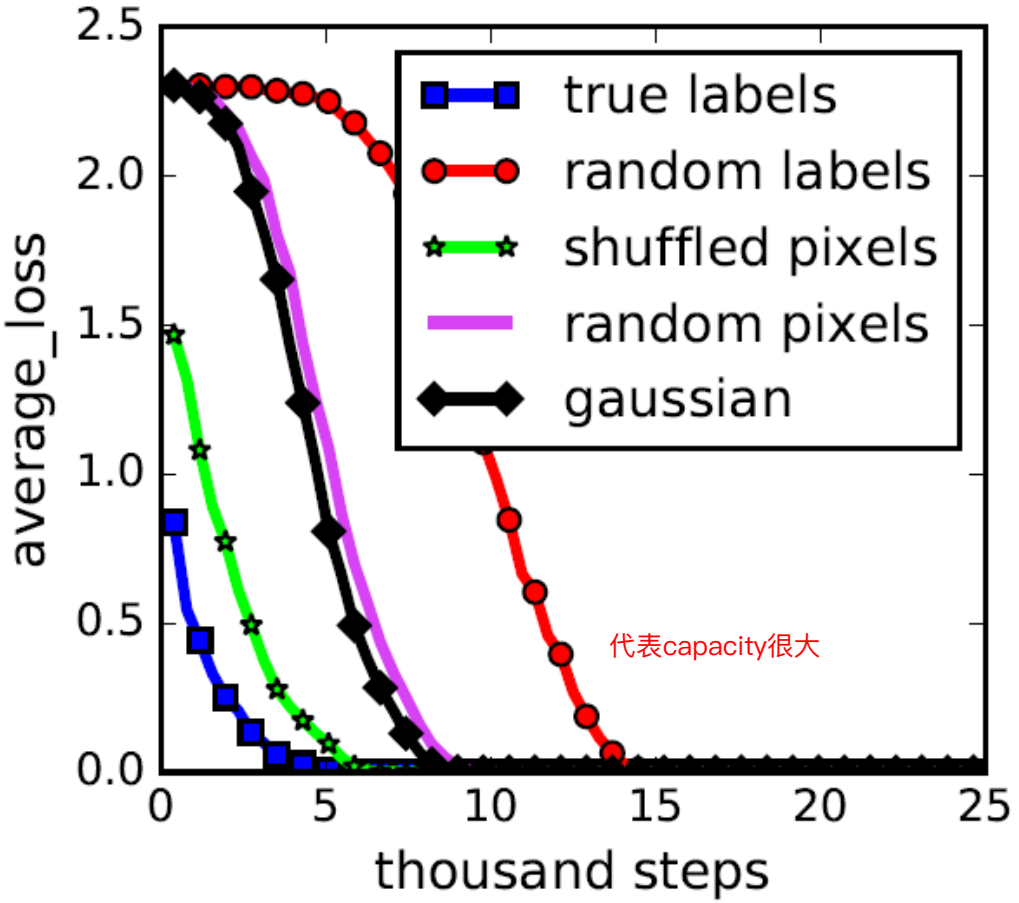
VC dimension (d_{VC}) of Linear Model < 4



Given 4 data points

What is the capacity of deep models?

Inception model on the CIFAR10



Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals, "Understanding deep learning requires rethinking generalization", ICLR 2017

Overparameterized Network?

No matter the data distribution

With probability $1 - \delta$

$$E_{test} \leq E_{train} + \Omega(R, M, \delta)$$

Smaller δ, larger Ω

R is the number of training data → Larger R, smaller Ω

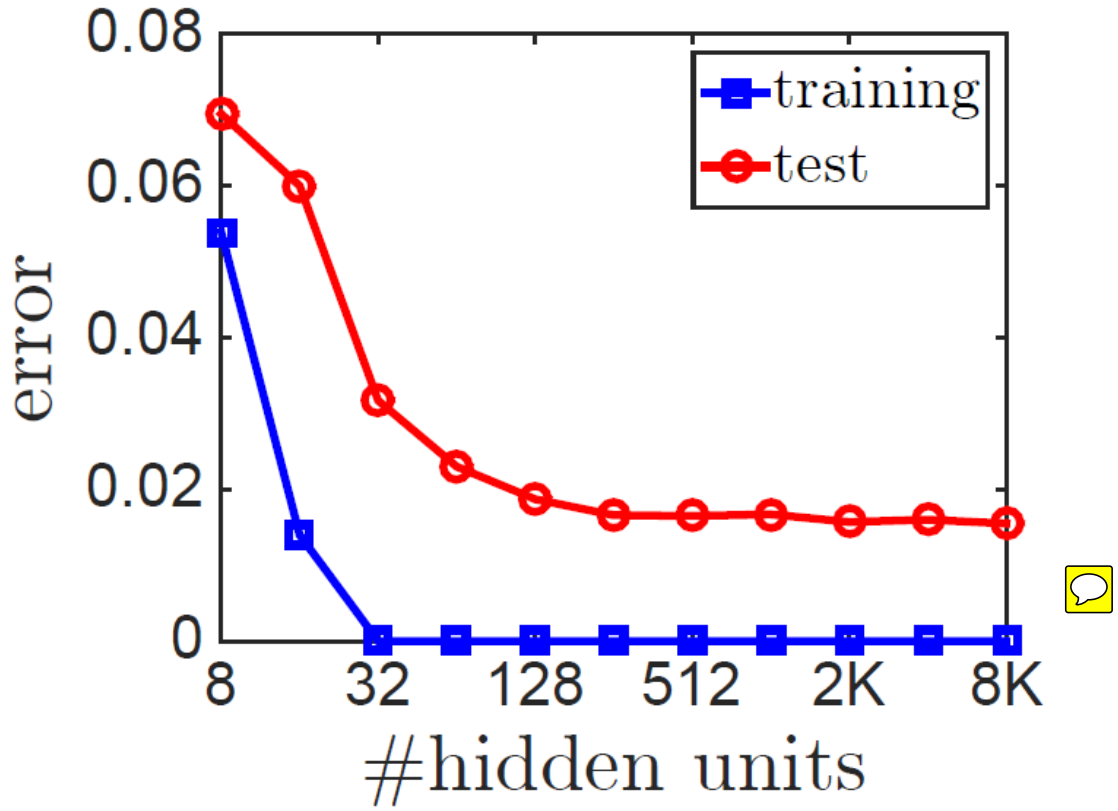
M is the "capacity" of your model → Larger M, larger Ω

("size" of the function set)

If two models have the same E_{train} → Select the one with smaller capacity

Demo

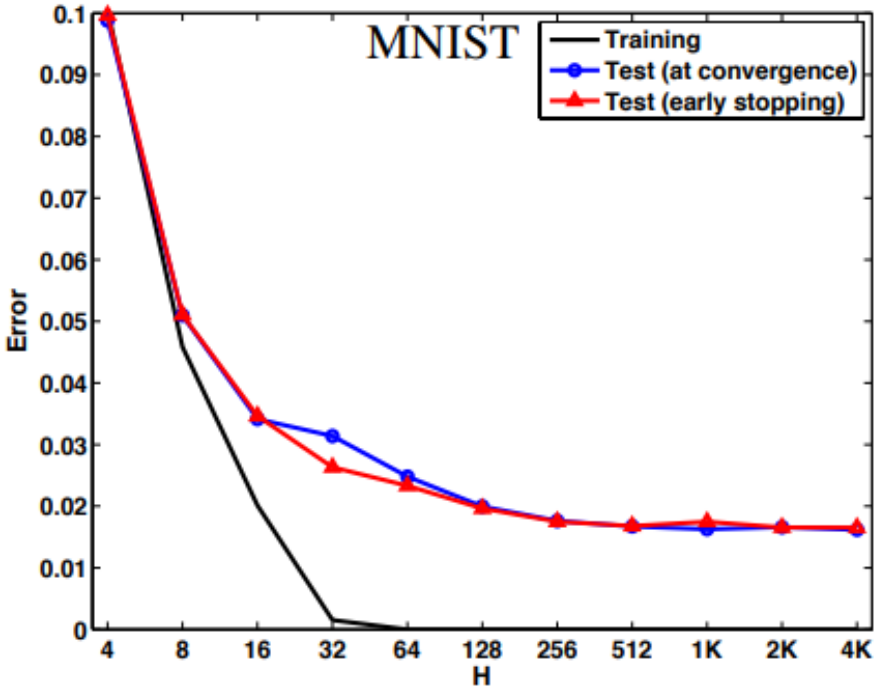
Overparameterized Network?

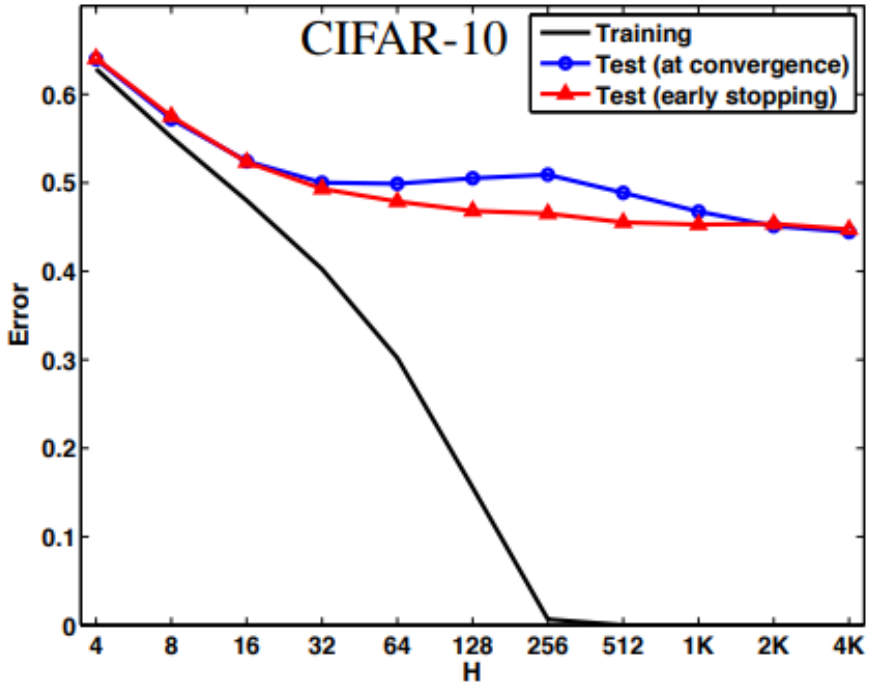


MNIST

https://arxiv.org/abs/1706.08947

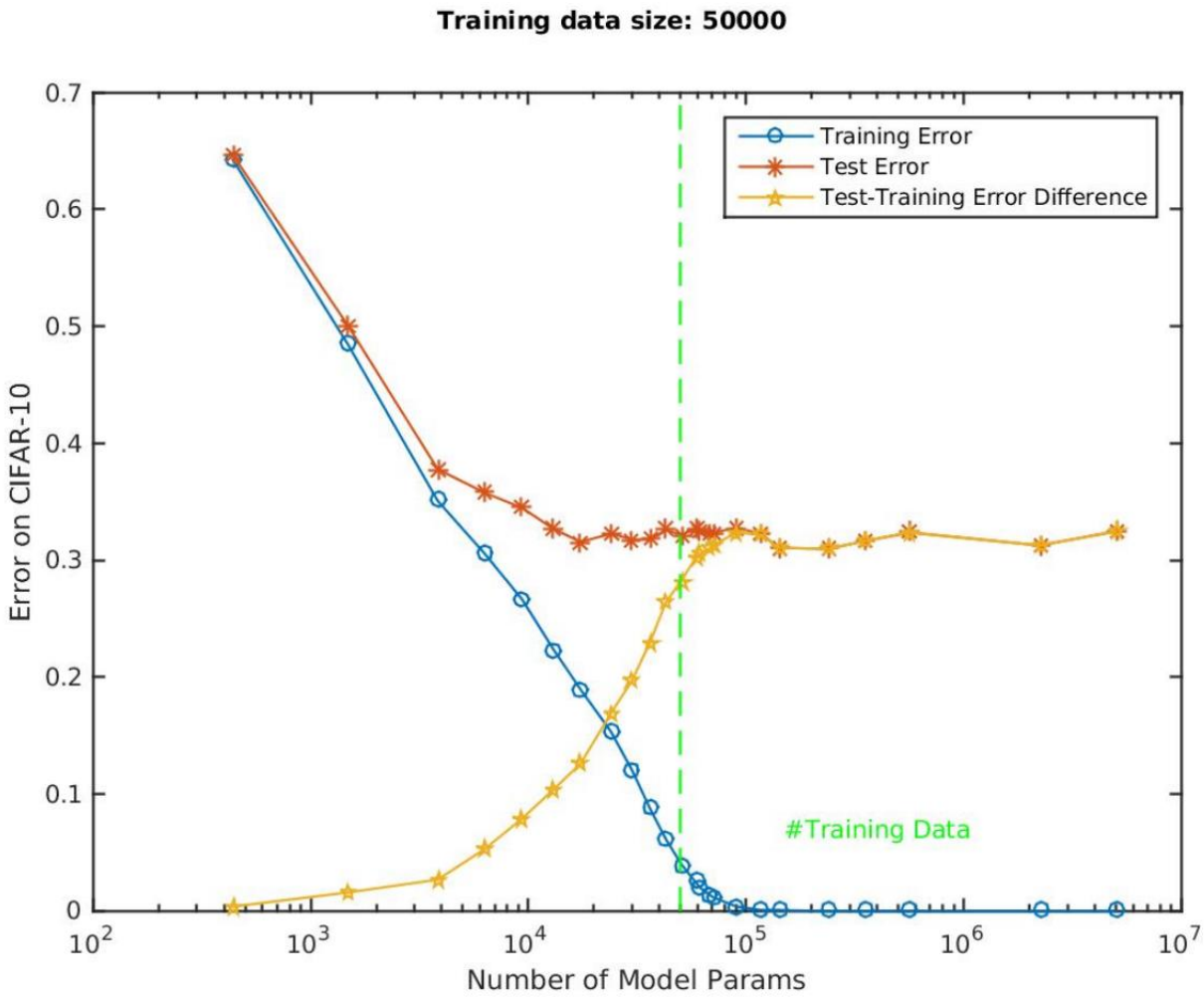
Overparameterized Network?



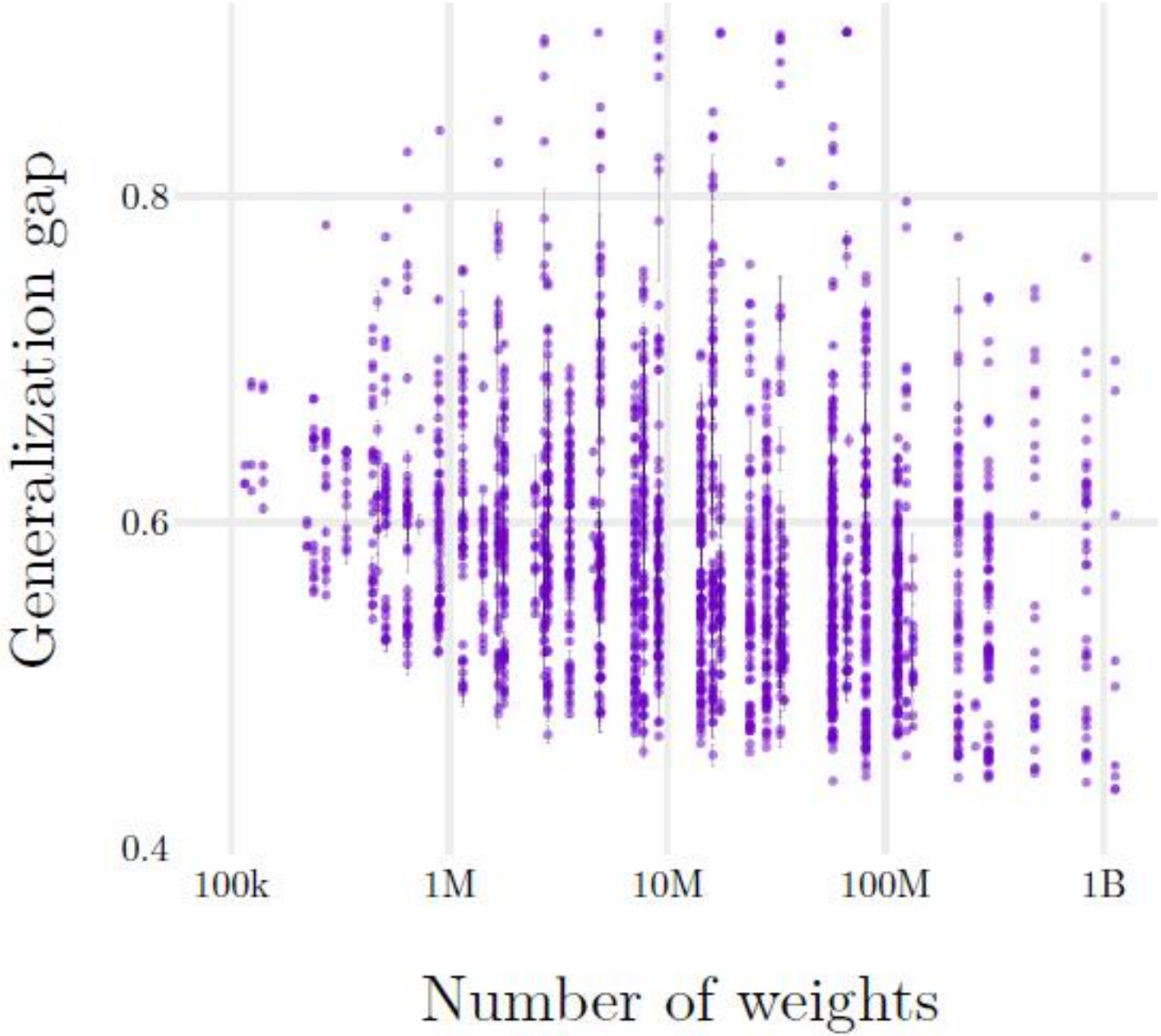


https://arxiv.org/pdf/1412.6614.pdf

Overparameterized Network?

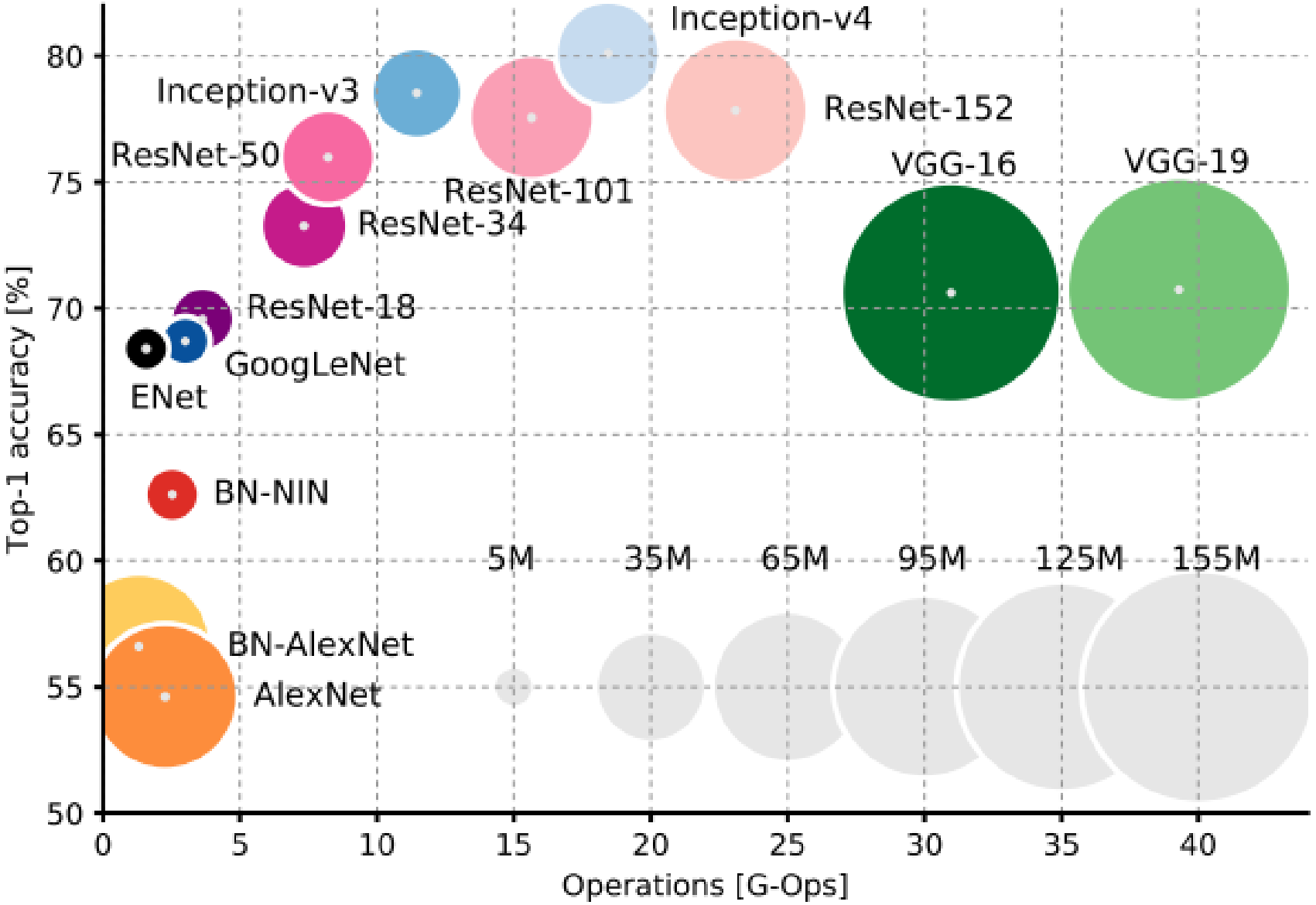


https://arxiv.org/abs/1801.00173



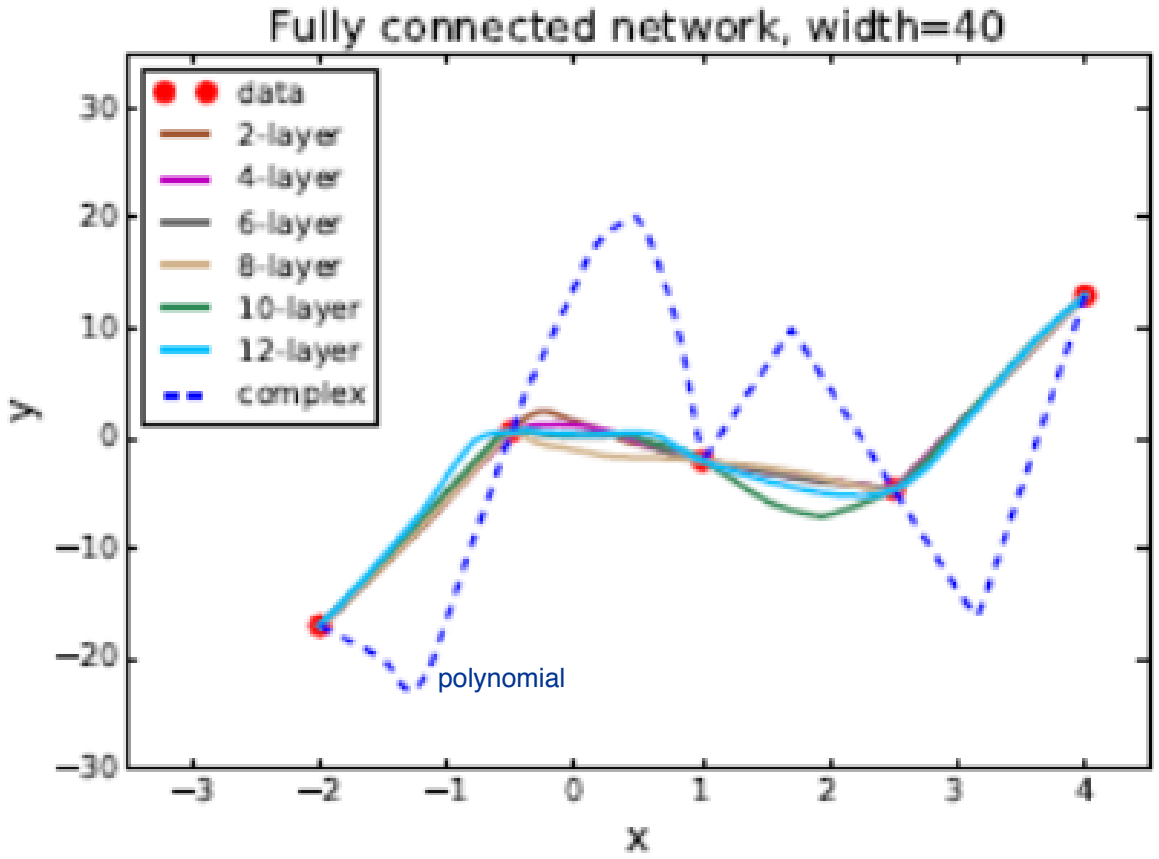
CIFIR-10, 100%
training accuracy

https://arxiv.org/pdf/1802.08760.pdf



https://arxiv.org/abs/1605.07678

Network regularizes itself?



https://arxiv.org/pdf/1706.10239.pdf

Concluding Remarks

- The capacity of deep model is large.
- However, it does not overfit!
- The reason is not clear yet.