SFU SIMON FRASER UNIVERSITY

McMaster University

Microsoft Research

Part IV:
Application of Deep Learning Model Complexity

Presenter: Jiang Bian

siam 2021 International Conference on Data Mining

How can model complexity analysis further benefit deep learning?

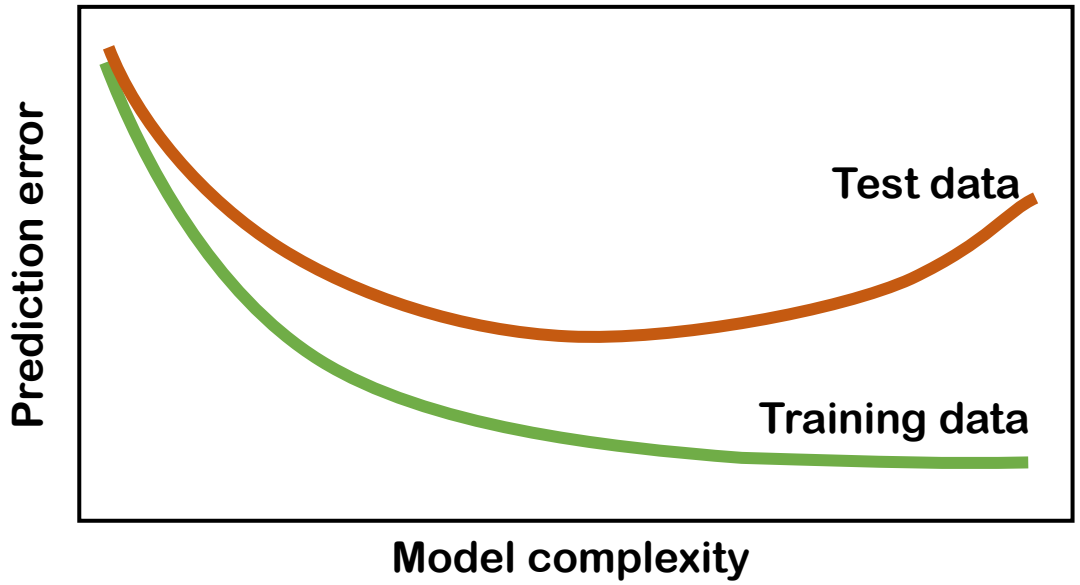
How can model complexity analysis further benefit deep learning?

Model Complexity in Generalization

"Machine learning is fundamentally about generalization."

[Mohri et. al.,2018]

- Model complexity and generalization are two closely related research directions.
- You must have seen this "Complexity and Generalization" figure:



Model Complexity in Generalization

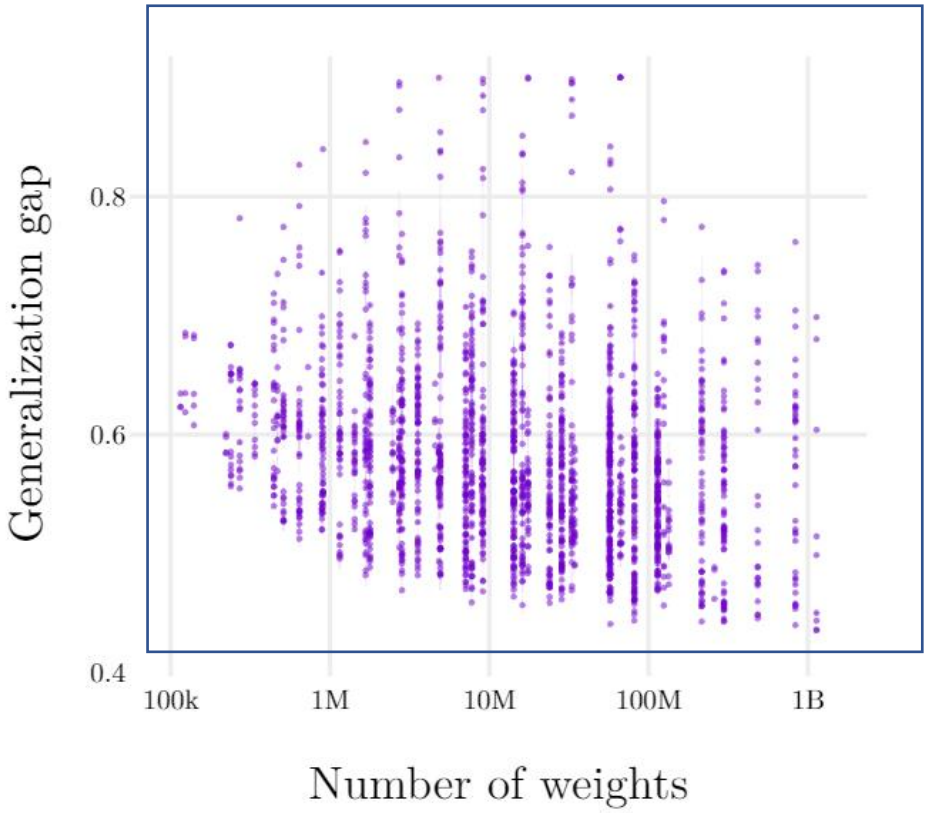


Figure from [Novak et. al., 2018]

But the situation seems different in deep learning

- Large and even over-parameterized neural networks exhibit good generalization ability
  - Sometimes, even better generalization compared to smaller or less complex deep models.

[neyshabur et al.,2017, zhang et al., 2017]

Model Complexity in Generalization

- Expressive capacity is used to bound generalization error

$$E(f_{A(D)}) - E_D(f_{A(D)}) \leq \sup_{f \in \mathcal{F}} \{E(f) - E_D(f)\}$$

[Kawaguchi et al., 2017]

Model Complexity in Generalization

Complexity measure that is expected to explain generalization should:

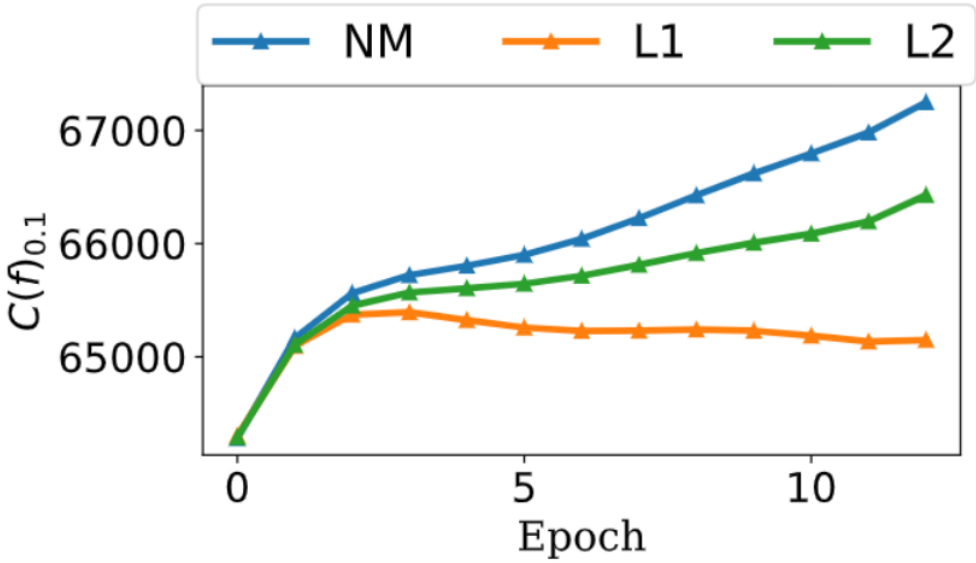
- With zero training error, a network trained on real labels, which leads to good generalization, is expected to have much lower complexity than a network trained on random labels.
- Increasing the number of hidden units or the number of parameters, which leads to a decreased generalization error, is expected to decrease the complexity measure.
- When training the same architecture on the same training dataset using two different optimization algorithms, if both lead to zero training errors, the model with better generalization is expected to have lower complexity.

[neyshabur et al.,2017]

How can model complexity analysis further benefit deep learning?

Model Complexity in Optimization

- Determine model parameters by minimizing a non-convex loss function.
- Provide a complexity-aware metric to make optimization traceable.
- Propose new regularization approaches based on complexity constraint.



Complexity measure during training of CIFAR.

[Hu et al., 2020]

Model Complexity in Optimization

- Inspire the explorations on the effectiveness of optimization approaches:

Are the advantages of deep learning related to optimization algorithms?

  - Deep, complex models might be easier to train than shallow architectures using the current optimization algorithms.

[Ba and Caruana, 2020]

How can model complexity analysis further benefit deep learning?

Complexity in Model Selection and Design

- How to determine a feasible model structure (type, size) for a specific learning task?
- How to pick a best one from a variety of models for a specific learning task?

Accuracy-Complexity Trade-off

High prediction accuracy is the key goal

An overly complex model may be difficult to train and incur unnecessary resource consumption

Complexity in Model Selection and Design

- How to determine a feasible model structure (type, size) for a specific learning task?
- How to pick a best one from a variety of models for a specific learning task?

Accuracy-Complexity Trade-off

High prediction accuracy is the key goal

An overly complex model may be difficult to train and incur unnecessary resource consumption

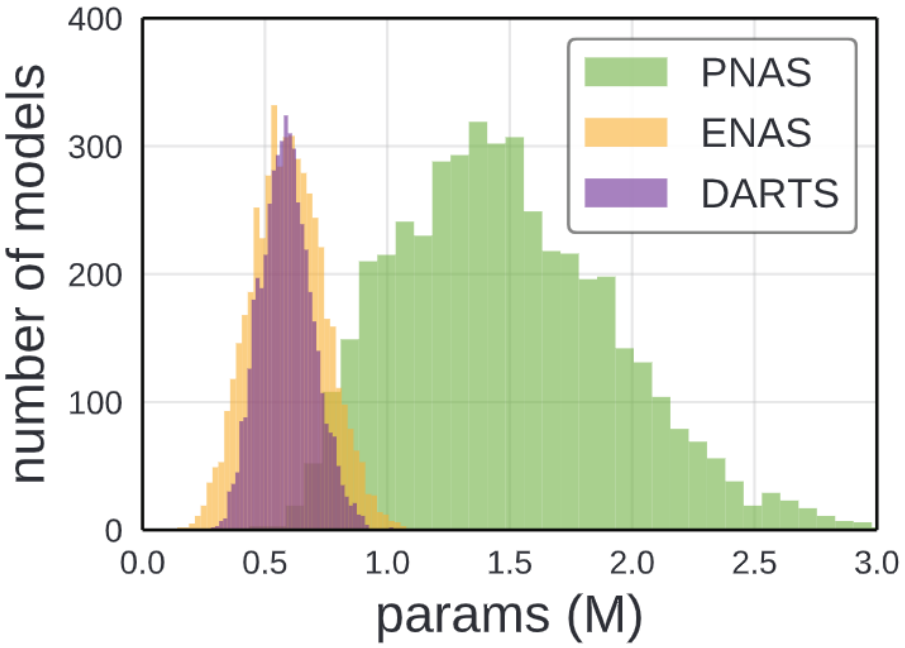
Pick a model with the “appropriate” complexity.

Complexity in Model Selection and Design

- Neural Architecture Search: automatically select a good neural network architecture for a given learning task
- Accuracy-complexity trade-off in NAS
  - Progressive Neural Architecture Search [Liu et al. 2018]: searches for convolutional neural network architectures in the increasing order of model complexity
  - Automatic Model Selection [Laredo et al., 2020]: searches for fully connected neural networks that yield a good balance between prediction accuracy and model complexity.

Complexity in Model Selection and Design

- Investigate the network design spaces



Complexity distribution of design spaces of several NAS model families.

[Radosavovic et al., 2019]