Landscape visualization and generalization

Beomgon Yu

Contents

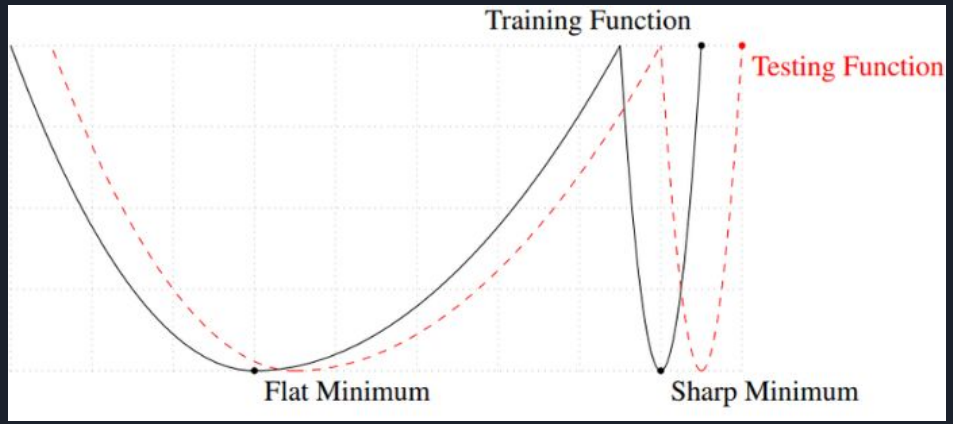
sharp vs flat

visualization

reference

sharp vs flat

Usually it is thought that flat minima has more generalization performance than sharp minima, but [1] questions about this assumption



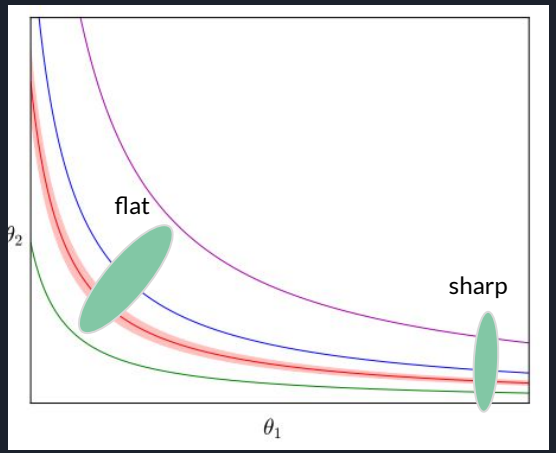
sharp vs flat

scale invariance

- conv - relu - conv와 같은 MLP 구조에서 first layer의 parameter에 10배를 곱하고 두번째 layer에서 1/10을 곱했을 때 output은 차이가 없다.

output이 동일하도록 scale을 변화시켜가며 그린 아래 그래프에서 flatness정도가 다른 두 weight에서 동일한 generalization error를 보인다.

flatness가 generalization에 관한 절대적인 지표는 아니지만 어느 정도 relation이 있는 지표로 생각됨



SGD vs adam

Usually adam get fast convergence, however SGD gets better generalization in some case, because adam falls into local minima and difficult to get out of it.

SGD is simple but adam is complicated (checking the history)

need to check optimizer beyond adam for solving poor generalization

batch size

small number of batch size have better generalization than large batch size

1. Before batchnorm, sampling of small number of batch size has various statistics, and batchnorm can study various input, but if batch size is big, randomness from sampling would be poor

2. suppose each sample has sharp minima.
   If batch size is big, this sharpness could be averaged to flat minima.

visualization

regarding [4]
concept of visualization is easy.
just sample some random point in weight space from any distribution.
and plot the loss between optimal point and sample point.

However, because of scale invariance, author of [4] say that it could be failed to capture intrinsic geometry of loss landscape.
(perturbation of loss is more sensitive when weight's norm is big)

author of [4] suggest filter-wise normalization by each layer and filter.

$$d_{i,j} \leftarrow \frac{d_{i,j}}{\|d_{i,j}\|}\|\theta_{i,j}\|$$

reference

1. https://arxiv.org/pdf/1703.04933.pdf
2. https://www.youtube.com/watch?v=5E9SFe5WU1s
3. https://www.inference.vc/sharp-vs-flat-minima-are-still-a-mystery-to-me/
4. https://proceedings.neurips.cc/paper/2018/file/a41b3bb3e6b050b6c9067c67f663b915-Paper.pdf
5. https://olaralex.com/visualizing-the-loss-landscape/
6. https://jithinjk.github.io/blog/nn_loss_visualized.md.html