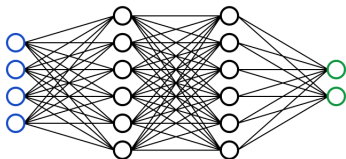


Introduction to Neural Networks

ML Instruction Team, Fall 2022

CE Department
Sharif University of Technology



TO DO

TO DO

TO DO

TO DO

TO DO

TO DO

TO DO

TO DO

TO DO

TO DO

Gradient Clipping

- In case of a large or small gradient, what will happen?

Gradient Clipping

- In case of a large or small gradient, what will happen?
- Gradient descent either **won't change our position** or will **send us far away**.

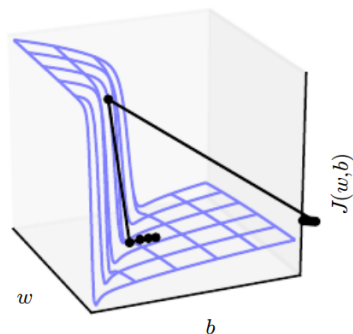


Figure: The problem of large gradient value [1].

Gradient Clipping

- Solve this problem simply by clipping gradient
- Clip the norm $\|g\|$ of the gradient g before updating parameters:

if $\|g\| > v$:

$$g \leftarrow \frac{g}{\|g\|} v$$

- v is the threshold for clipping which is a hyperparameter
- Gradient clipping saves the direction of gradient and controls its norm

Gradient Clipping

- The effect of gradient clipping:

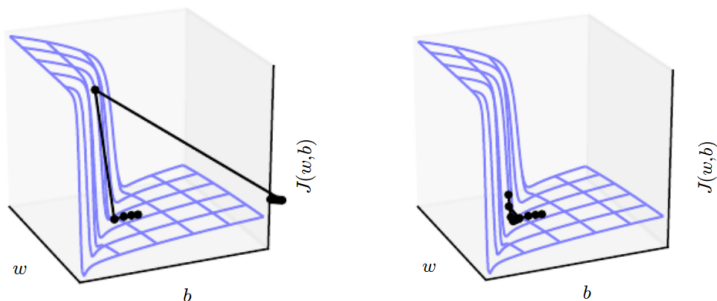


Figure: The "cliffs" landscape (left) without gradient clipping and (right) with gradient clipping [1].

Weight Initialization

Various GD types

Final Notes

Thank You!

Any Question?

References



I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*.
MIT Press, 2016.
<http://www.deeplearningbook.org>.