Notes on Deep Neural Networks

Clay L. McLeod — clay.l.mcleod@gmail.com

January 24, 2016

1 Neural Network Basics

1.1 Weight Initializations

1.1.1 Lecun's Distribution

Presented in [1, Sec 4.6].

$$X \sim N\left[-\frac{1}{\sqrt{n_j}}, \frac{1}{\sqrt{n_j}} \right] \tag{1}$$

$$X \sim U\left[-\frac{3}{\sqrt{n_j}}, \frac{3}{\sqrt{n_j}} \right] \tag{2}$$

1.1.2 Glorot Distribution

Presented in [2, 4.2].

$$X \sim N\left[-\frac{\sqrt{2}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{2}}{\sqrt{n_j + n_{j+1}}} \right] \tag{3}$$

$$X \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}} \right] \tag{4}$$

1.1.3 He Distribution

Previous assumes linear activation function, which is not suitable for ReLU or its derivatives. Thus, the authors of [3][pg. 4] derive a theoretically sound initialization for networks utilizing the ReLU activation family.

$$X \sim N \left[\; -\sqrt{\frac{2}{n_j}}, \; \sqrt{\frac{2}{n_j}} \; \right] \tag{5}$$

$$X \sim U \left[\; -\sqrt{\frac{6}{n_j}}, \; \sqrt{\frac{6}{n_j}} \; \right] \tag{6}$$

References

[1] Y. A. LeCun, L. Bottou, G. B. Orr, and K.-R. Müller, "Efficient backprop," in *Neural networks: Tricks of the trade*, pp. 9–48, Springer, 2012.

[2] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *International conference on artificial intelligence and statistics*, pp. 249–256, 2010.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015.