

ImageNet Classification with Deep Convolutional Neural Networks

Xuwen Yang

May 5. 2018

The standard way to model a neurons output f as a function of its input x is with (1) or(2)

$$f(x) = \tanh(x) \quad (1)$$

$$f(x) = (1 + e^{-x})^{-1} \quad (2)$$

In terms of training time with gradient descent, these saturating nonlinearities are much slower than the non-saturating nonlinearity $f(x) = \max(0; x)$. Following Nair and Hinton [1], we refer to neurons with this nonlinearity as Rectified Linear Units (ReLUs). Deep convolutional neural networks with ReLUs train several times faster than their equivalents with tanh units. This is demonstrated in Figure 1.

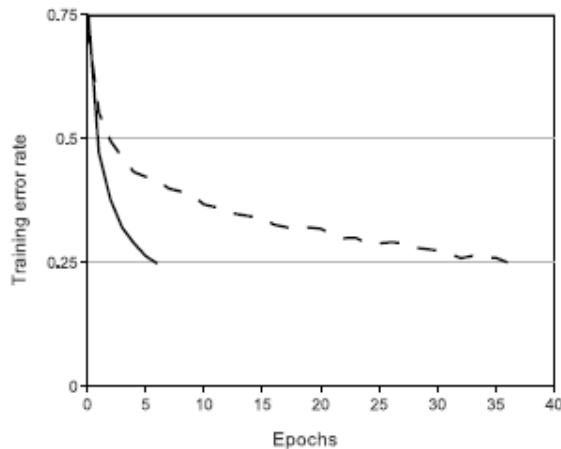


Figure 1: A four-layer convolutional neural network with ReLUs (solid line) reaches a 25% training error rate on CIFAR-10 six times faster than an equivalent network with tanh neurons

which shows the number of iterations required to reach 25% training error on the CIFAR-10 dataset for a particular four-layer convolutional network. This plot shows that we would not have been able to experiment with such large neural networks for this work if we had used traditional saturating neuron models.

References

- [1] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *International Conference on International Conference on Machine Learning*, pages 807–814, 2010.