Deep Sparse Rectifier Neural Networks, Date 5/9/17 No.

Glorot, Bordes, Bengio

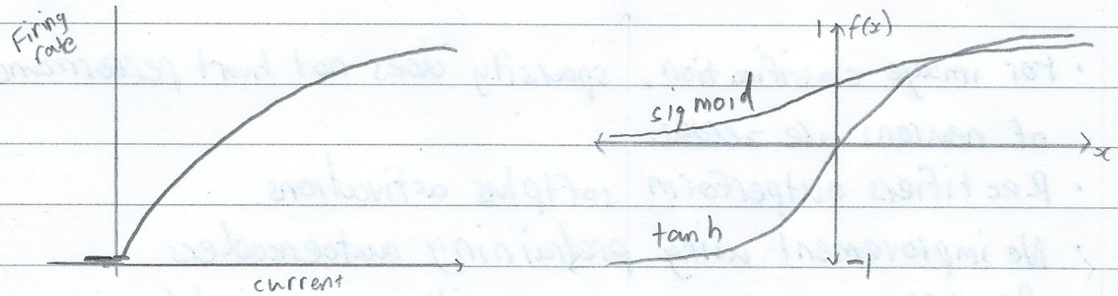
The rectifier function max(0,x) is both a useful model for neuron activation in neuroscience, and an efficient activation function in neural networks.

$\nearrow 1 \to -1$

- Biological neurons can have activations that are asymmetric,

$1 \to 1 \leftarrow$ symmetric, one-sided. $\leftarrow$ $1 \to 0$

- Neurons encode info in a sparse manner: only 1-4% are active at once. Tradeoff between expressiveness and low energy use. Neural nets without L1 reg do not have this property
- Biological neurons use very different activation functions to NNs.



- tanh is preferred to sigmoid because its steady state is zero
- however, its asymmetry about zero is not present in biology.

Sparsity has a number of computational benefits:

- information disentangling - robust to small input changes
- efficient variable-size representation - inputs can be in a variable-size data structure
- more likely to be linearly separable

However, excess sparsity may reduce predictive capability.

Rectifier Networks

- Because real neurons rarely reach saturation, they can be well approximated by a rectifier.
- A rectifier automatically produces sparsity: 50% will be initialised to zero.
- Much easier to compute than tanh.
- The softplus function $\log(1+e^x)$ can avoid hard zeroes, but the zeroes seem to be good for NNs.
- Rectifier nets need more hidden units, to represent any antisymmetry in the data.

Experimental results

- For image classification, sparsity does not hurt performance until 85% of neurons are zeroes.
- Rectifiers outperform softplus activations.
- No improvement using pretraining autoencoders
- Rectifiers work very well with supervised / semi-supervised problems, but in the latter case pre-training is needed.
- Very strong performance on text sentiment analysis: lower RMSE than tanh, at 50% sparsity.