Computer Science Department, University of Bonn

Knowledge Graph Analysis

Exercise Sheet 6

Dr. Hamed Shariat Yazdi, Prof. Jens Lehmann

December 11, 2018

1 In Class

1. **Transfer function**
   Consider the sigmoid transfer function

$$\sigma(u) = \frac{1}{1 + e^{-u}} \ ,$$

   Show that its derivative fulfills

$$\sigma'(u) = \big(1 - \sigma(u)\big) \cdot \sigma(u) \ .$$

2. **Weight parameters**
   Consider a neuron in a hidden layer with membrane potential u, and firing rate ν. Let $x = (x_1, \ldots, x_m) \in \mathbb{R}^p$ denote the input, let $w = (w_1, \ldots, w_p)$ denote the connection weights from the inputs to a hidden node, and let b denote the bias weight:

$$u = \sum_{i=1}^{p} w_i \cdot x_i + b = w^T x + b$$
$$\nu = \sigma(u)$$

   Say we want to make the sigmoid σ twice as steep, ideally we'd like to apply

$$\tau(u) = \frac{1}{1 + e^{-2u}} \ ,$$

instead of σ. But we cannot do this, since the sigmoid itself is fixed; it does not have any parameters. How can the parameters (weights and bias) be changed to achieve the same effect? Let w, b denote the parameters used with τ, how to set w', b' used with σ to model the same function?

3. **Size of the hidden layer**
   Consider two neural networks, one with a small and one with a huge hidden layer, trained to minimize the empirical risk. For which one would you expect higher/lower

   ▷ training error

   ▷ over-fitting

   ▷ generalization error?

   For which reasons?