COMPUTER SCIENCE DEPARTMENT, UNIVERSITY OF BONN

Knowledge Graph Analysis

Solutions to Exercise Sheet 6

Dr. Hamed Shariat Yazdi, Prof. Jens Lehmann

December 20, 2018

1 In Class

1. **Transfer function**
   We know that

$$(\frac{f}{g})' = \frac{f'g - fg'}{g^2} \quad .$$

   With this we obtain

$$\sigma'(u) = \frac{e^{-u}}{(1+e^{-u})^2} = \frac{e^{-u}}{1+e^{-u}}.\frac{1}{1+e^{-u}} = \frac{e^{-u}}{1+e^{-u}}.\sigma(u) \quad .$$

   We can see that

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

   And thus

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

2. **Weight parameters**
   It holds $\tau(u) = \sigma(2 \cdot u)$, hence the problem is solved by doubling the input to the transfer function. We aim for

$$\sum_{i=1} w_i x_i + b = 2 \cdot \left[\sum_{i=1} w_i' x_i + b' \right] .$$

This is easily realized with the choice $w'_i = 2 \cdot w_i$ and $b' = 2 \cdot b$. Hence doubling all weights and biases allows us "make the sigmoid twice as steep".

3. **Size of the hidden layer**
   The larger network will tend to have a lower training error because it can use the additional degrees of freedom to improve the fit on the training data. This automatically increases the risk of over-fitting. The two effects have a contrarious effect on the generalization error, hence in general we cannot know anything about which network generalizes better.