Проблема затухающих градиентов

Vanishing gradients problem

Человек, на голове которого была шляпа с павлиньим пером, зашел в бар

Vanishing gradients problem

$$\frac{\partial E_t}{\partial W} = \sum_{k=0}^{t} \frac{\partial E_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial s_t} \frac{\partial s_t}{\partial s_k} \frac{\partial s_k}{\partial W}$$

Vanishing gradients problem

$$\frac{\partial E_t}{\partial W} = \sum_{k=0}^{t} \frac{\partial E_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial s_t} \frac{\partial s_t}{\partial s_k} \frac{\partial s_k}{\partial W}$$

$$\prod_{j=k+1}^{t} \frac{\partial s_j}{\partial s_{j-1}}$$

Vanishing gradients problem

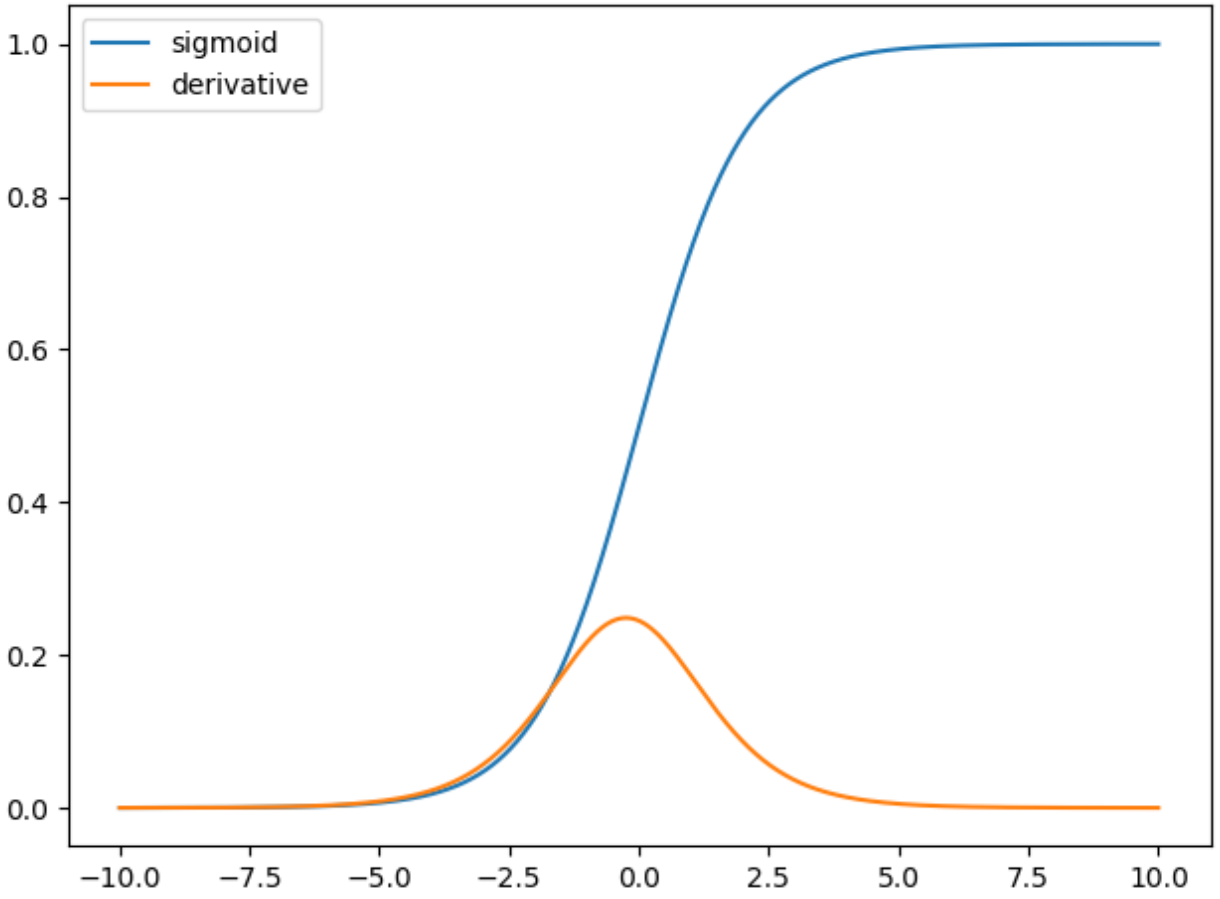
$$\frac{\partial E_t}{\partial W} = \sum_{k=0}^{t} \frac{\partial E_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial s_t} \left(\prod_{j=k+1}^{t} \frac{\partial s_j}{\partial s_{j-1}} \right) \frac{\partial s_k}{\partial W}$$

Vanishing gradients problem

$$\frac{\partial E_t}{\partial W} = \sum_{k=0}^{t} \frac{\partial E_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial s_t} \left(\prod_{j=k+1}^{t} \frac{\partial s_j}{\partial s_{j-1}} \right) \frac{\partial s_k}{\partial W}$$

$$W \frac{\partial f}{\partial (W s_{t-1})}$$

Sigmoid and derivative



Vanishing gradients problem

$$\frac{\partial E_t}{\partial W} = \sum_{k=0}^{t} \frac{\partial E_t}{\partial \hat{y}_t} \frac{\partial \hat{y}_t}{\partial s_t} \left(\prod_{j=k+1}^{t} \frac{\partial s_j}{\partial s_{j-1}} \right) \frac{\partial s_k}{\partial W}$$

Vanishing gradients problem

Человек, на голове которого была шляпа с павлиньим пером, зашел в бар

$$\frac{\partial E_{10}}{\partial W} = \frac{\partial E_{10}}{\partial \hat{y}_{10}} \frac{\partial \hat{y}_{10}}{\partial s_{10}} \left(\prod_{j=1}^{10} \frac{\partial s_j}{\partial s_{j-1}} \right) \frac{\partial s_1}{\partial W} + \dots$$

Как бороться?

- Другая функция активации (хотя в RNN чаще используется tanh и sigmoid)

- LSTM ?