

# Тятя! Тятя! Нейросети заменили продавца!

Ппилиф Ульяновкин

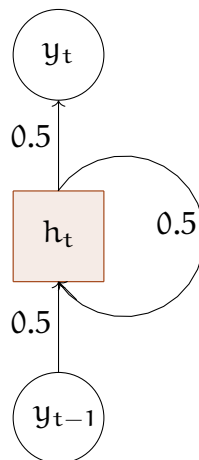
[https://github.com/FUlyankin/neural\\_nets\\_prob](https://github.com/FUlyankin/neural_nets_prob)

## Листочек 7: рекуррентные сети

В ожидании чудес, невозможных чудес  
Я смотрю в темноту, но я не верю в прогресс  
Я хочу быть убийцей, я хочу быть живым  
Мне осталось всего лишь дожить до седин  
*Нейронная Оборона (нейросеть, обученная на стихах Летова)*

### Упражнение 1 (Туда и обратно)

Маша хочет сделать шаг обратного распространения ошибки через рекуррентную ячейку для последовательности  $y_0 = 0, y_1 = 1, y_2 = -1, y_3 = 2$ . Скрытое состояние инициализировано как  $h_0 = 0$ . Все веса инициализированы как 0.5. Во всех уравнениях, описывающих ячейку нет констант. В качестве функций активаций Маша использует ReLU. В качестве функции потерь Маша использует MSE.



- Сделайте прямой шаг через ячейку. Для каждого элемента последовательности постройте прогноз. Посчитайте значение ошибки.
- Сделайте обратный шаг распространения ошибки. Посчитайте для каждого из весов градиенты и обновите значения весов.

## Упражнение 2 (Число параметров)

У Маши есть очень длинный временной ряд. Она хочет обучить несколько нейросетей предсказывать его дальнейшее значение. В своих моделях Маша нигде не использует константы.

- а. Маша выделяет окно длины 100. Оно движется по последовательности. Для каждого окна Маша предсказывает следующее значение в ряду. В сетку подаются наблюдения с 1-го по 100-е. Прогнозируется 101-ое наблюдение. Затем на вход подаются наблюдения со 2-го по 100-е. Прогнозируется 102-ое наблюдение. И так далее до конца последовательности. На первом слое используется 20 нейронов. На втором слое используется один нейрон. Сколько параметров нужно оценить?
- б. Маша использует одну простую RNN-ячейку. Сколько параметров ей необходимо оценить?
- в. Маша хочет предсказывать значение  $y_t$  по трём последовательностям  $y_{t-1}$ ,  $y_{t-2}$  и  $y_{t-3}$ . На первом слое сети Маша использует два рекуррентных нейрона. На втором слое она использует один рекуррентный нейрон. Матрица какого размера идёт на вход в первый слой? Матрица какого размера передаётся во второй слой? Какое число параметров необходимо оценить Маше?
- г. Мы находимся в условиях прошлого пункта, но используем LSTM-ячейки с забыванием. Сколько параметров надо оценить?
- д. Мы находимся в условиях прошлого пункта, но используем GRU-ячейки. Сколько параметров надо оценить?

## Упражнение 3 (Из картинки в формулу)

У Маши есть два рекуррентных нейрона. Помогите ей изобразить их в виде вычислительных графов.

Однонаправленный:

$$h_t = f_h(b_h + W \cdot h_{t-1} + V \cdot x_t)$$

$$y_t = f_y(b_y + U \cdot h_t)$$

Двунаправленный:

$$h_t = f_h(b_h + W \cdot h_{t-1} + V \cdot x_t)$$

$$s_t = f_s(b_s + W' \cdot s_{t+1} + V' \cdot x_t)$$

$$o_t = b_y + U \cdot h_t + U' \cdot s_t$$

$$y_t = f_y(o_t)$$

## Упражнение 4 (Замочные скважины)

В 2000 году Шмидхубер и Герс предложили модификацию LSTM с замочными скважинами. Она описывается следующей системой из уравнений

$$\begin{aligned}c'_t &= \phi_c(W_c x_t + V_c h_{t-1} + b_c) \\ i_t &= \phi_i(W_i x_t + V_i h_{t-1} + U_i c_{t-1} + b_i) \\ f_t &= \phi_f(W_f x_t + V_f h_{t-1} + U_f c_{t-1} + b_f) \\ o_t &= \phi_o(W_o x_t + V_o h_{t-1} + U_o c_{t-1} + b_o) \\ c_t &= f_t \odot c_{t-1} + i_t \odot c'_t \\ h_t &= o_t \odot \phi_h(c_t)\end{aligned}$$

Изобразите эту ячейку в виде вычислительного графа. Объясните, чем именно она отличается от базовой модификации LSTM. Какой в этом смысл?

## Упражнение 5 (Лишние части)

Выпишите уравнения, описывающие LSTM-ячейку с забыванием и GRU-ячейку. Какие последовательности и веса нужно занулить, чтобы эти ячейки превратились в простую RNN-ячейку?