

Long short-term memory

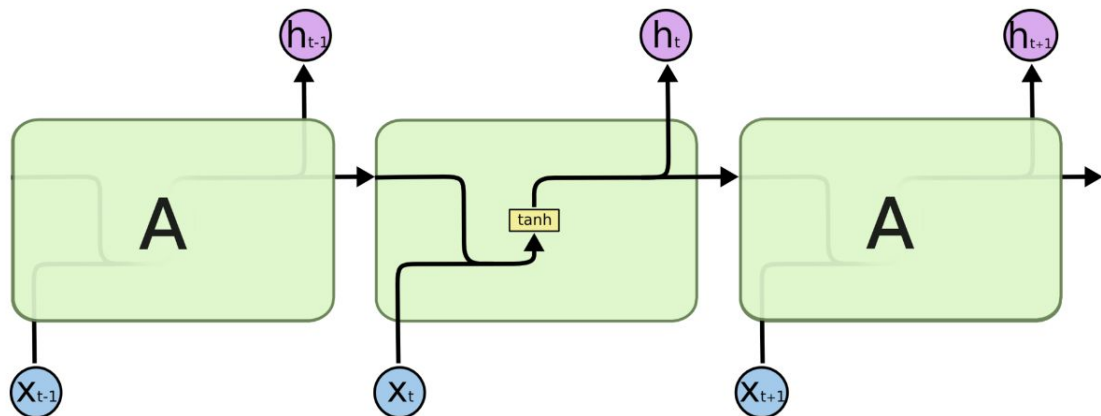
Применение

- автодополнение кода
- машинный перевод
- аннотация картинок
- создание музыки
- генерация человеческого почерка

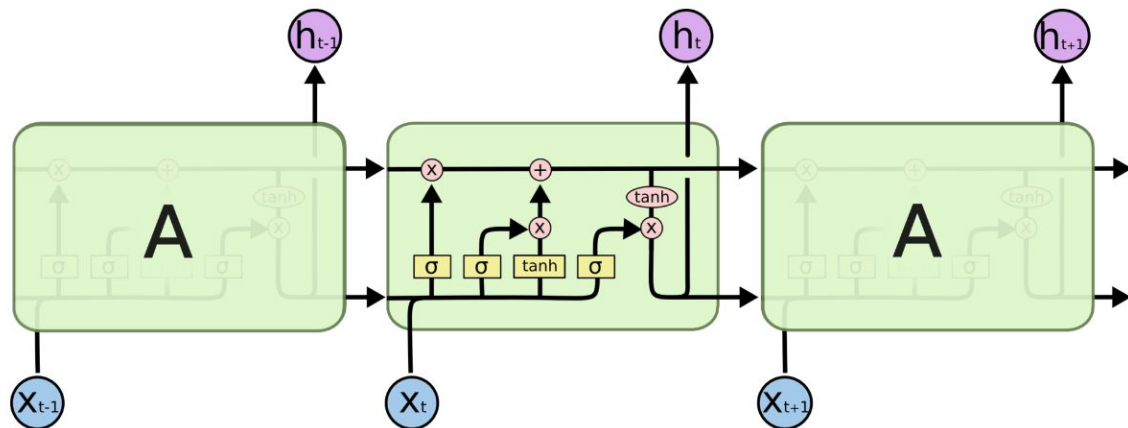
Hello, Darya!

Hello Darya!

Hello, Darya!



Повторяющийся модуль в стандартной RNN состоит из одного слоя.



Повторяющийся модель в LSTM сети состоит из четырех взаимодействующих слоев.

Обозначения



Copy



Concatenate



Vector
Transfer

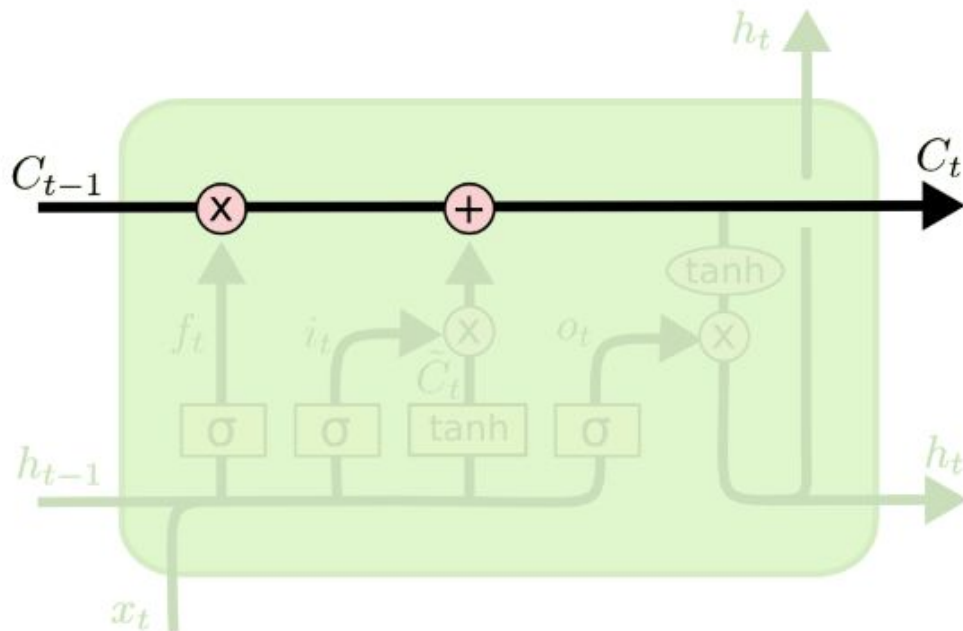


Pointwise
Operation



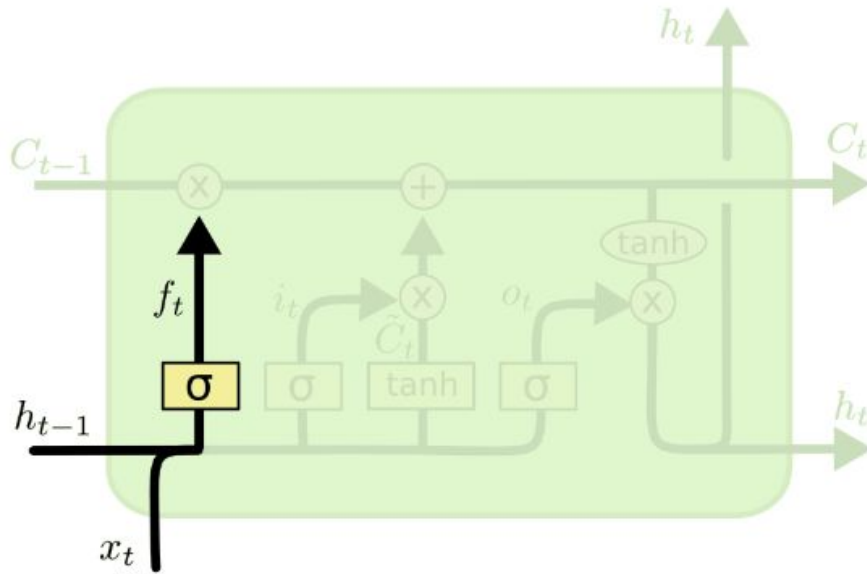
Neural Network
Layer

Основная идея LSTM



- “Вентиль забывания”
- “Вентиль добавления”
- “Вентиль внимания”

“Вентиль забывания” (forget gate)



$$f_t = \sigma(W_r x_t + U_r h_{t-1}), \sigma \in [0, 1]$$

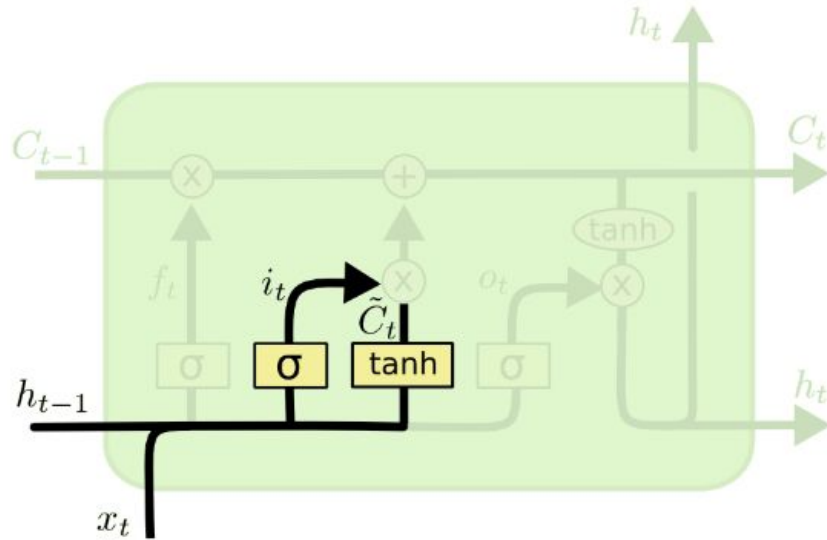
f_t - вентиль забывания

c_t - состояние ячейки
(долговременная память)

h_t - рабочая память
(скрытое состояние)

x_t - входные данные

“Вентиль добавления” (save gate layer)



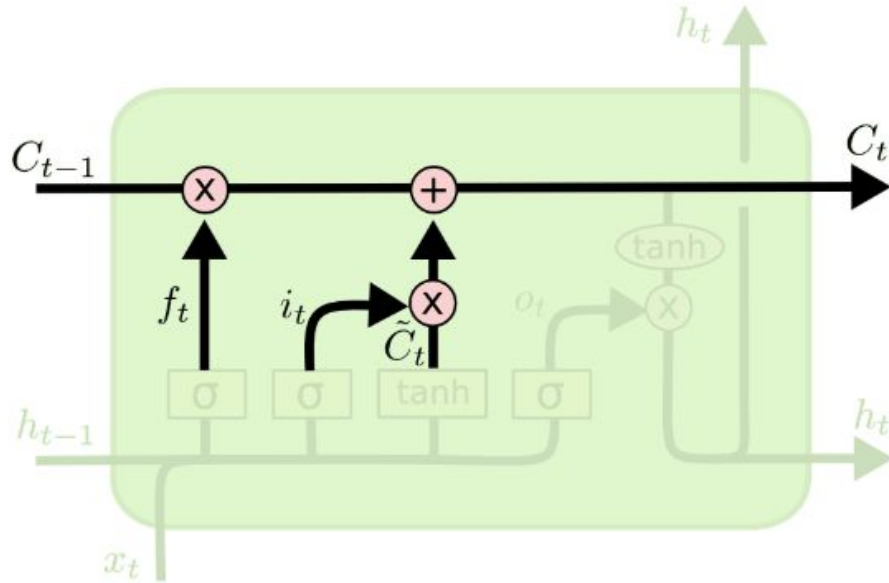
$$i_t = \sigma(W_s x_t + U_s h_{t-1})$$

$$\hat{c}_t = \tanh(W_l x_t + U_l h_{t-1})$$

i_t - вентиль добавления

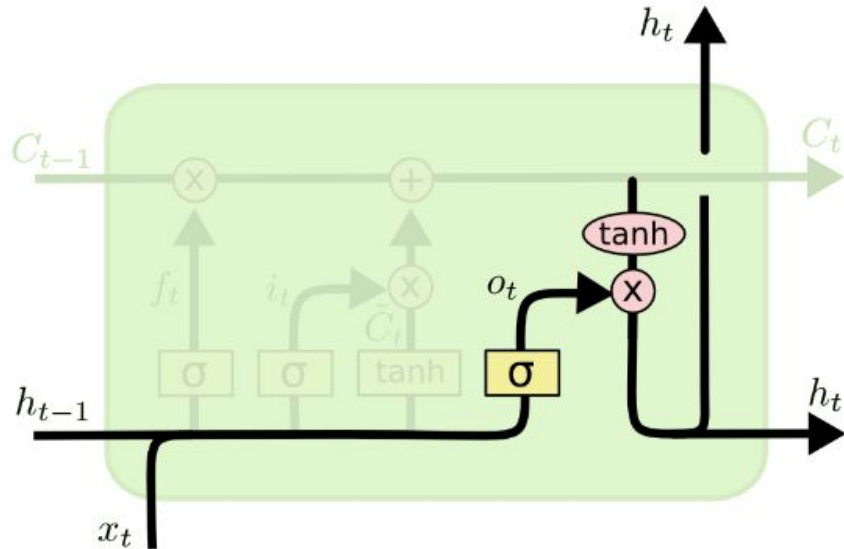
\hat{c}_t - функция активации

Обновление состояния ячейки



$$c_t = f_t \cdot c_{t-1} + i_t \cdot \hat{c}_t$$

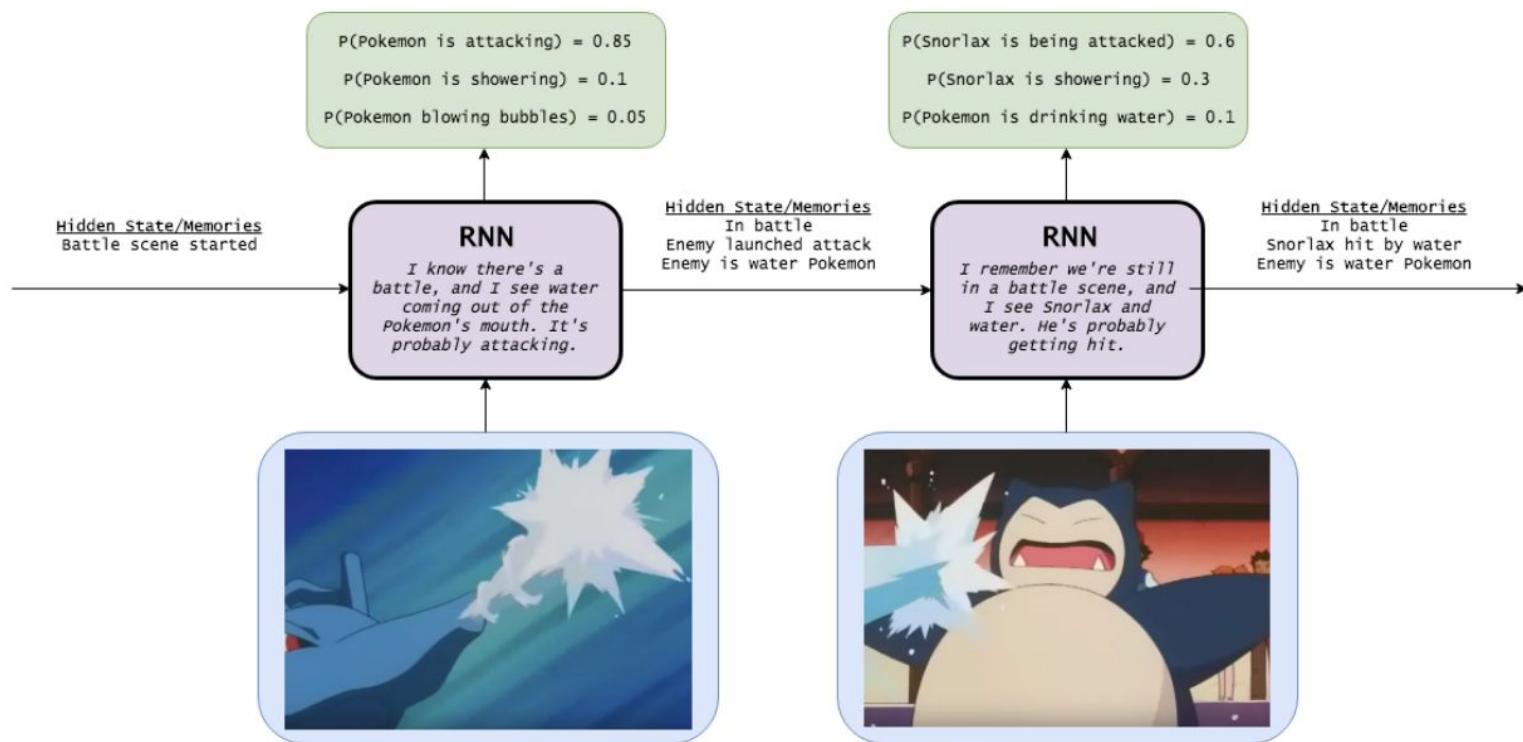
“Вентиль внимания” (attention gate layer)

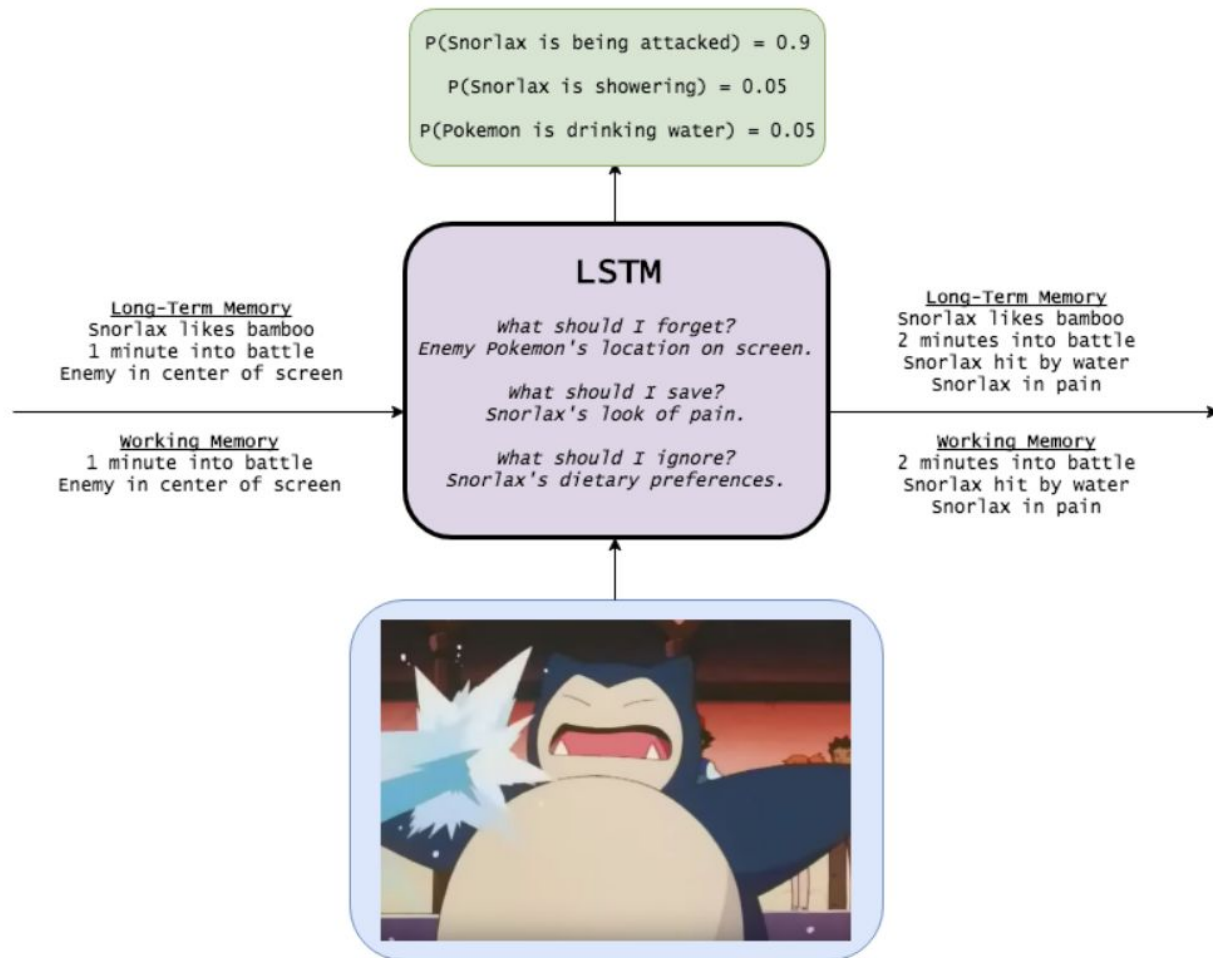


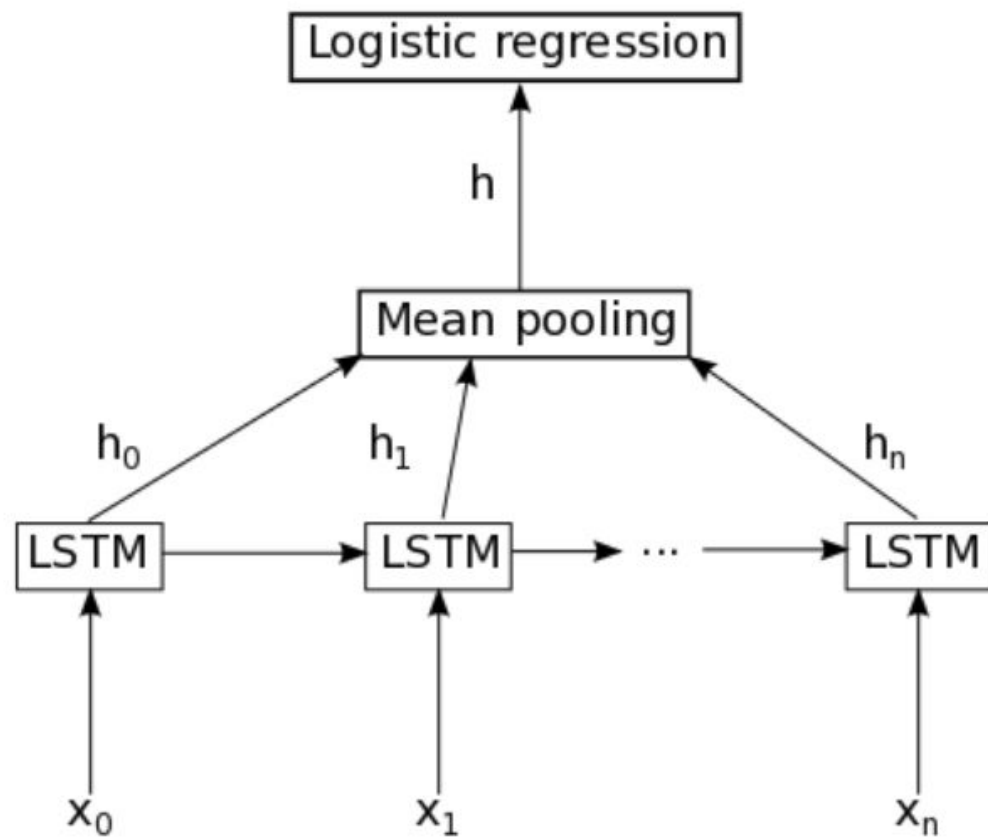
$$o_t = \sigma(W_f x_t + U_f h_{t-1})$$

$$h_t = o_t \cdot \tanh(c_t)$$

o_t - ВЕНТИЛЬ ВНИМАНИЯ







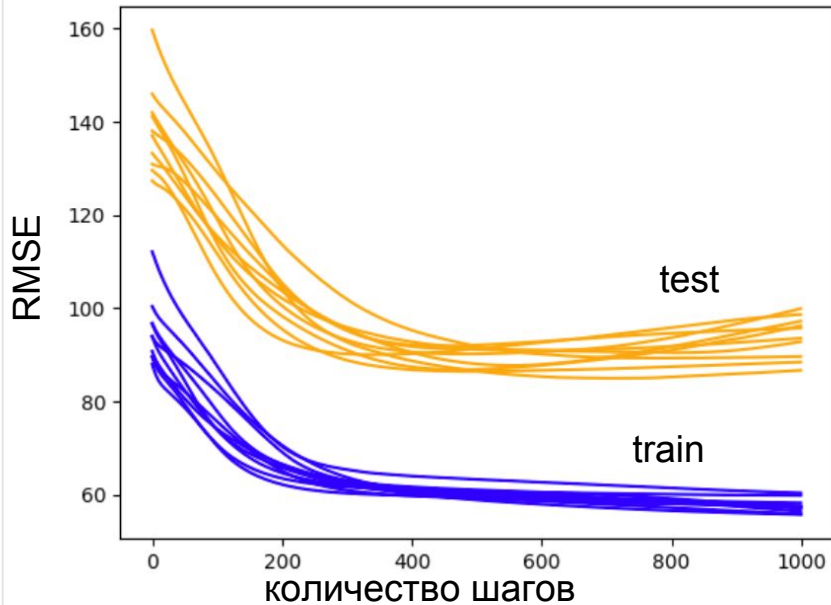
Почему LSTM спасает от взрыва градиентов

$$\frac{\partial c_t}{\partial c_{t-1}} = \frac{\partial f_t \cdot c_{t-1}}{\partial c_{t-1}} = c_{t-1} \cdot \frac{\partial f_t}{\partial c_{t-1}} + f_t \cdot \frac{\partial c_{t-1}}{\partial c_{t-1}} = 0 + 1 \cdot f_t$$

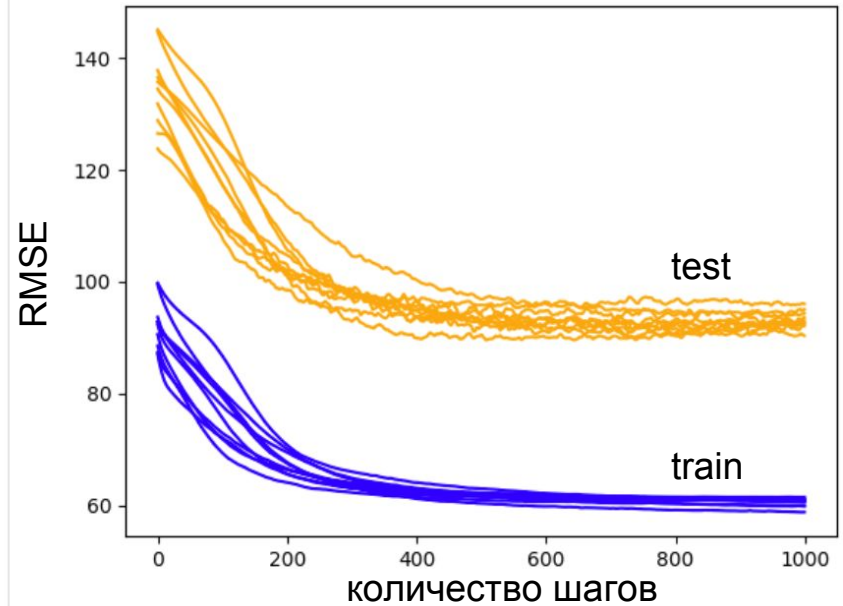
$$\frac{\partial c_{t_1}}{\partial c_t} = \prod_{k=1}^{t_1-t} f_{t+k}$$

$$f_t = \sigma(W_r x_t + U_r h_{t-1}), \sigma \in [0, 1]$$

Dropout



Diagnostic Line Plot of the Baseline Model on the Shampoo Sales Daset



Diagnostic Line Plot of Recurrent Dropout Performance on the Shampoo Sales Dataset

Нормализация данных

- Преобразовать временной ряд в стационарный
- Масштабировать данные
- Преобразовать данные в данные для обучения с учителем

Недостатки

- LSTM - сеть неэффективна по памяти.
- Переобученность, несмотря на дропауты, все равно остается.
- Нужен большой датасет для обучения.

Источники

- <http://datareview.info/article/issleduem-lstm-seti-chast-2/>
- <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- <http://datareview.info/article/issleduem-lstm-seti-chast-1/>
- <https://habrahabr.ru/company/wunderfund/blog/331310/>
- <http://www.cs.toronto.edu/~graves/handwriting.cgi?text=Hello%2C+Darya%21&style=&bias=0.15&samples=3>
- <http://deeplearning.net/tutorial/lstm.html>
- <https://d-nb.info/1082034037/34>
- <https://machinelearningmastery.com/use-dropout-lstm-networks-time-series-forecasting/>