



SCHOOL
OF ENERGY
& POWER ENGINEERING

Архитектуры нейронных сетей для обработки последовательностей

Сергей Владимирович Аксёнов,

Доцент отделения информационных технологий ИШИТР,

Томский политехнический университет

Томск-2023

Рекуррентные нейронные сети

Вид архитектур, где связи между элементами образуют направленную последовательность.

Применяются при работе с:



Сигналами с датчиков;



Распознаванием рукописного текста;



Анализом речи;



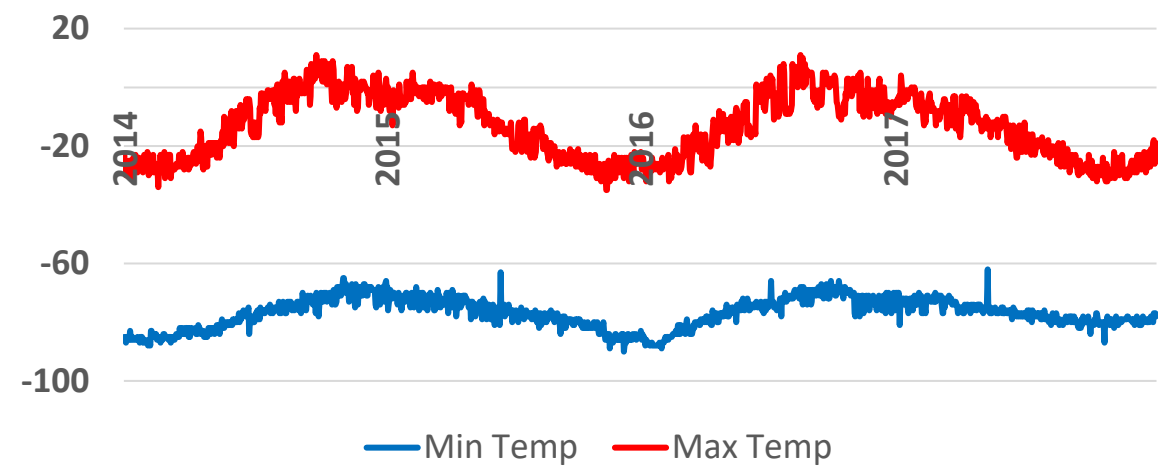
Обработкой кадров видео;



и т.д.

Обработка последовательностей

Сигналы



Дата	Мин. темп, °C	Макс. темп, °C	Давлен., Pa
25-8-2016	-74	10	828
24-8-2016	-70	4	828
23-8-2016	-71	5	828
22-8-2016	-68	11	823
21-8-2016	-71	7	821
20-8-2016	-72	4	825
19-8-2016	-71	4	823

Текст

Эпидемиологическое прогнозирование заболеваемости клещевыми инфекциями на основе биотических и абиотических факторов с помощью нейросетей.

Подготовка данных

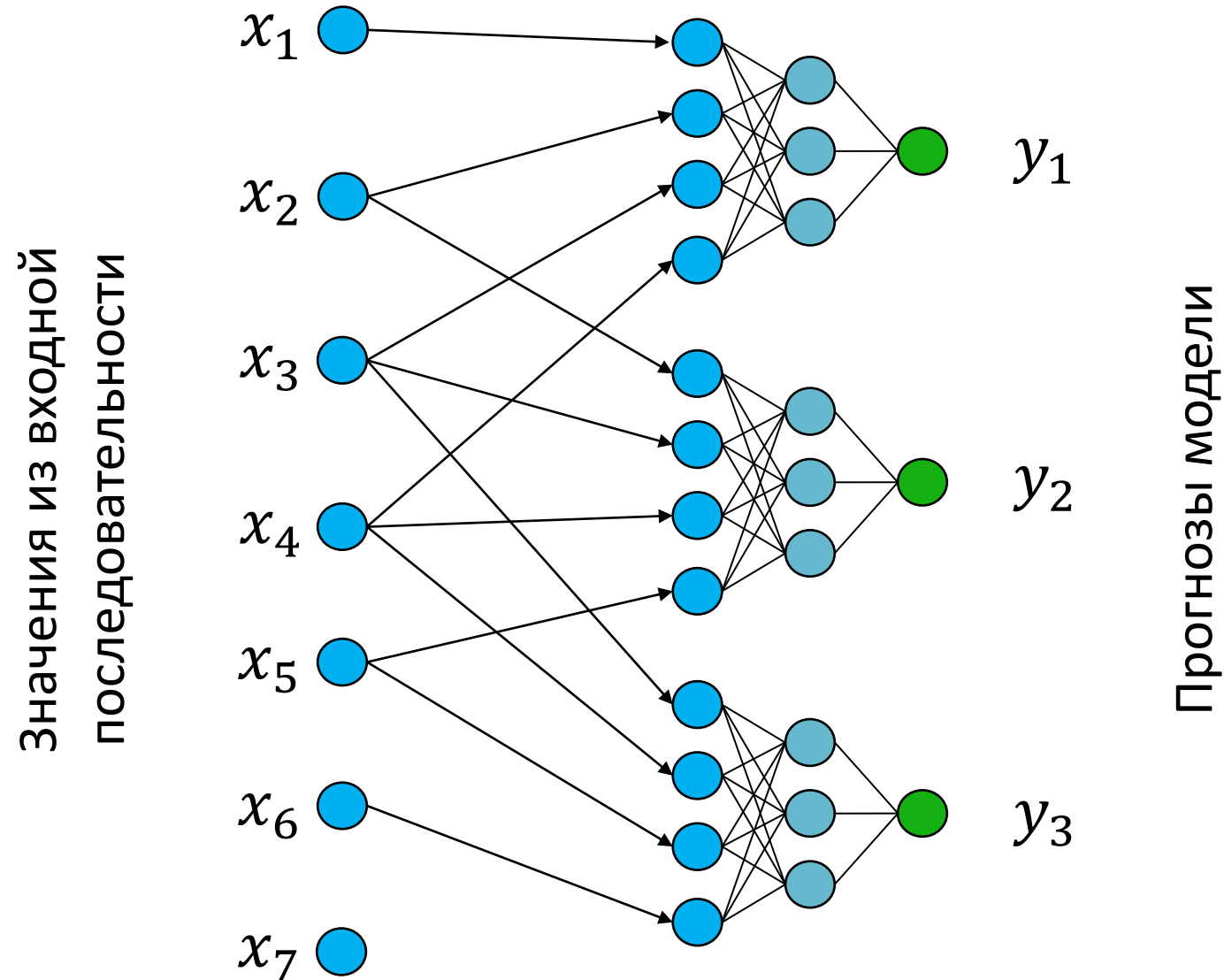
Дата	Макс. темп
10-8-2016	-6
11-8-2016	8
12-8-2016	4
13-8-2016	4
14-8-2016	5
15-8-2016	7
16-8-2016	7
17-8-2016	6
18-8-2016	0
19-8-2016	4
20-8-2016	4
21-8-2016	7
22-8-2016	11
23-8-2016	5



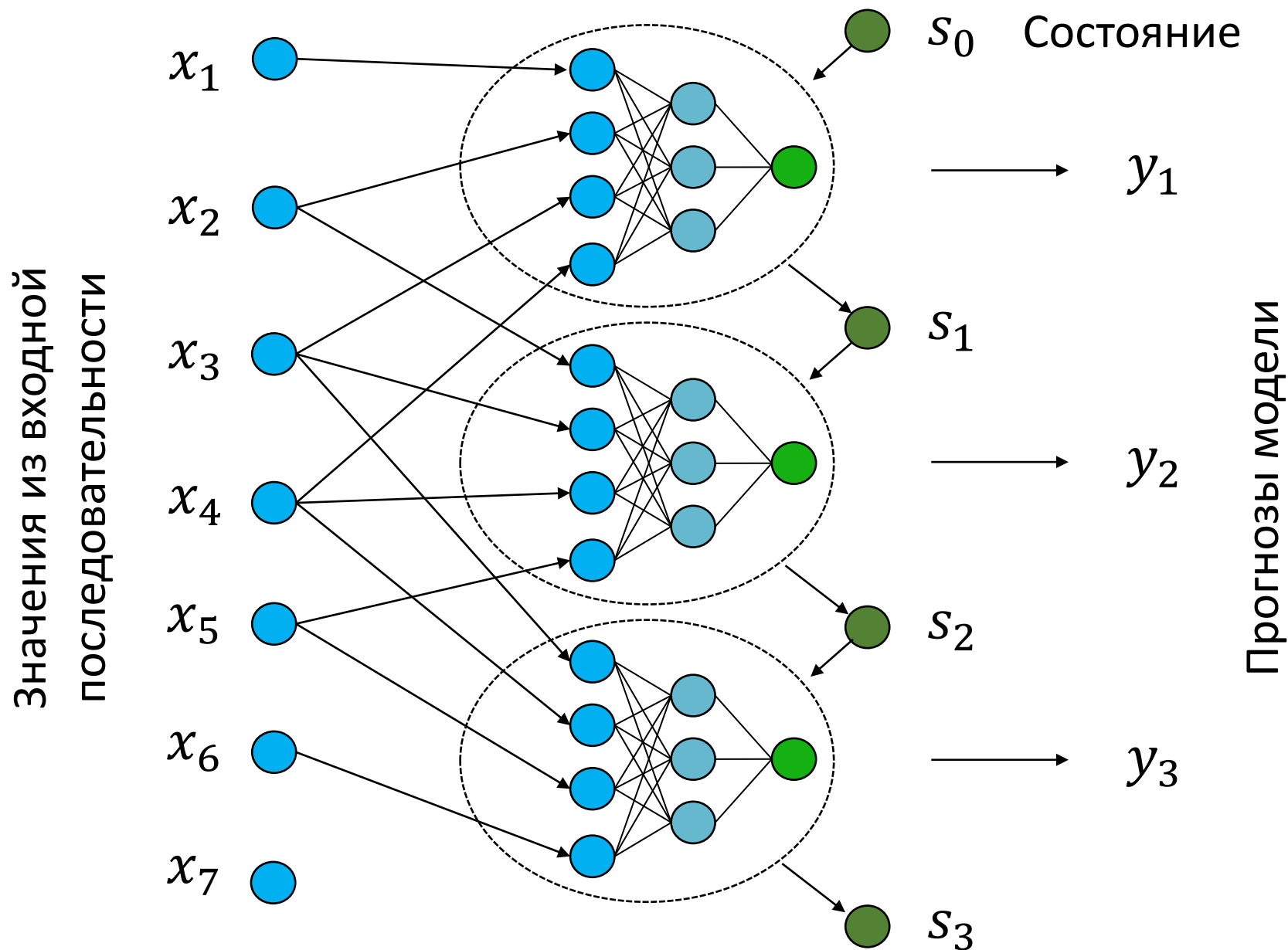
Новая таблица для числа дней до прогноза = 6

D-6	D-5	D-4	D-3	D-2	D-1	Прогноз (D)
-6	8	4	4	5	7	7
8	4	4	5	7	7	6
4	4	5	7	7	6	0
4	5	7	7	6	0	4
5	7	7	6	0	4	4
7	7	6	0	4	4	7
7	6	0	4	4	7	11
6	0	4	4	7	11	5

Использование сети прямого распространения

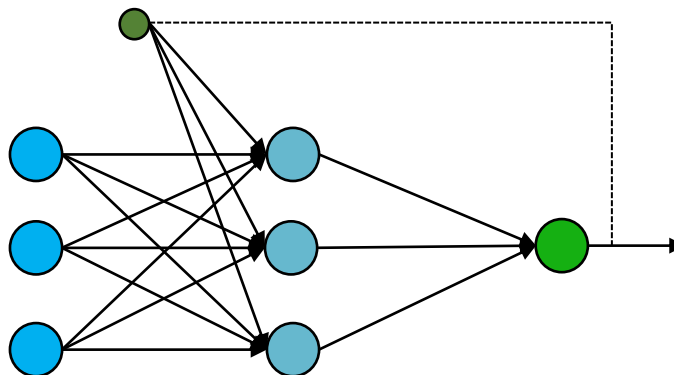


Использование предыдущего состояния

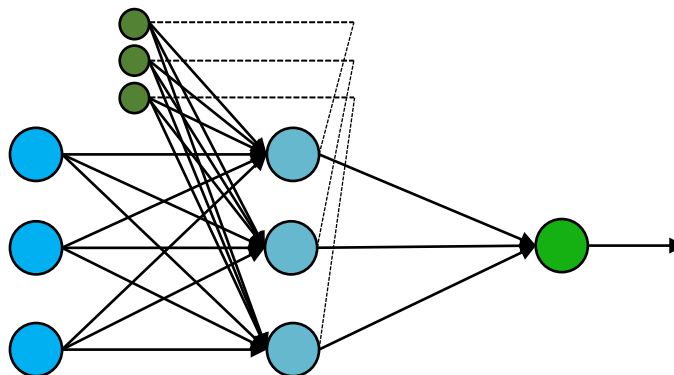


Примеры рекуррентных моделей

Сеть Джордана



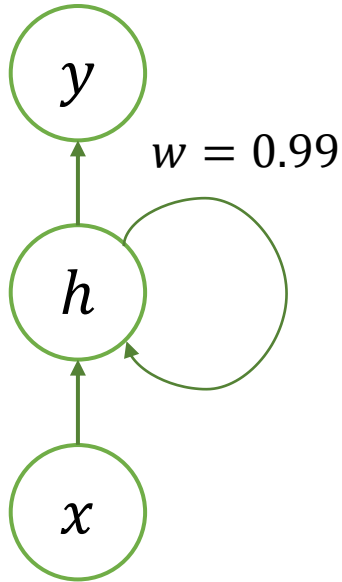
Сеть Эльмана



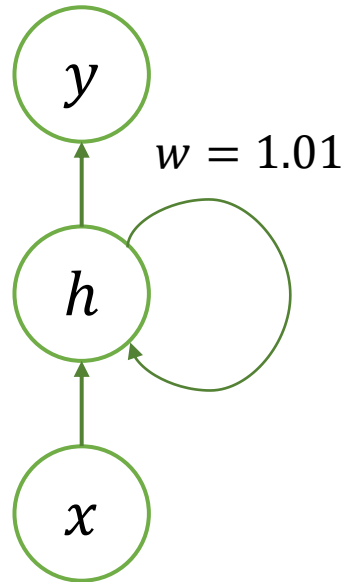
Затухающий и взрывающийся градиент

Затухающий градиент: Взрывающийся градиент: Желаемый градиент:

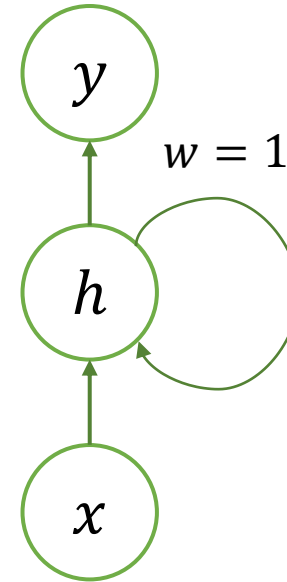
$$|w_{hh}| < 1$$



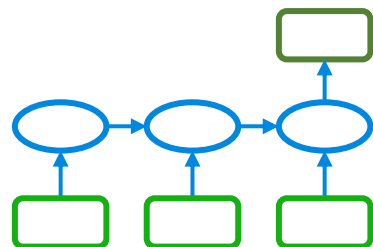
$$|w_{hh}| > 1$$



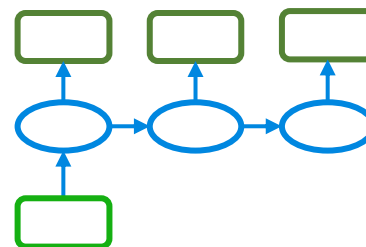
$$|w_{hh}| = 1$$



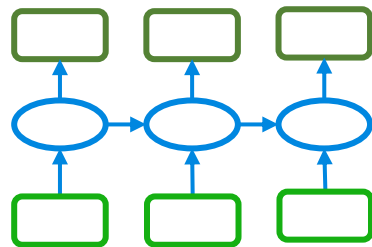
Задачи с последовательностями



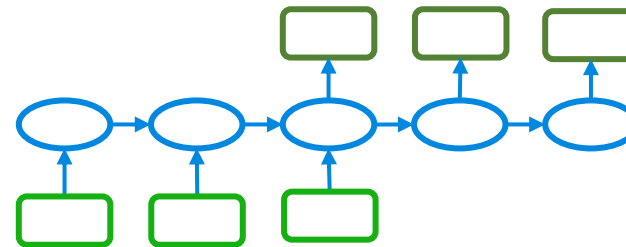
Последовательность
входов и один выход



Один вход и
последовательность выходов

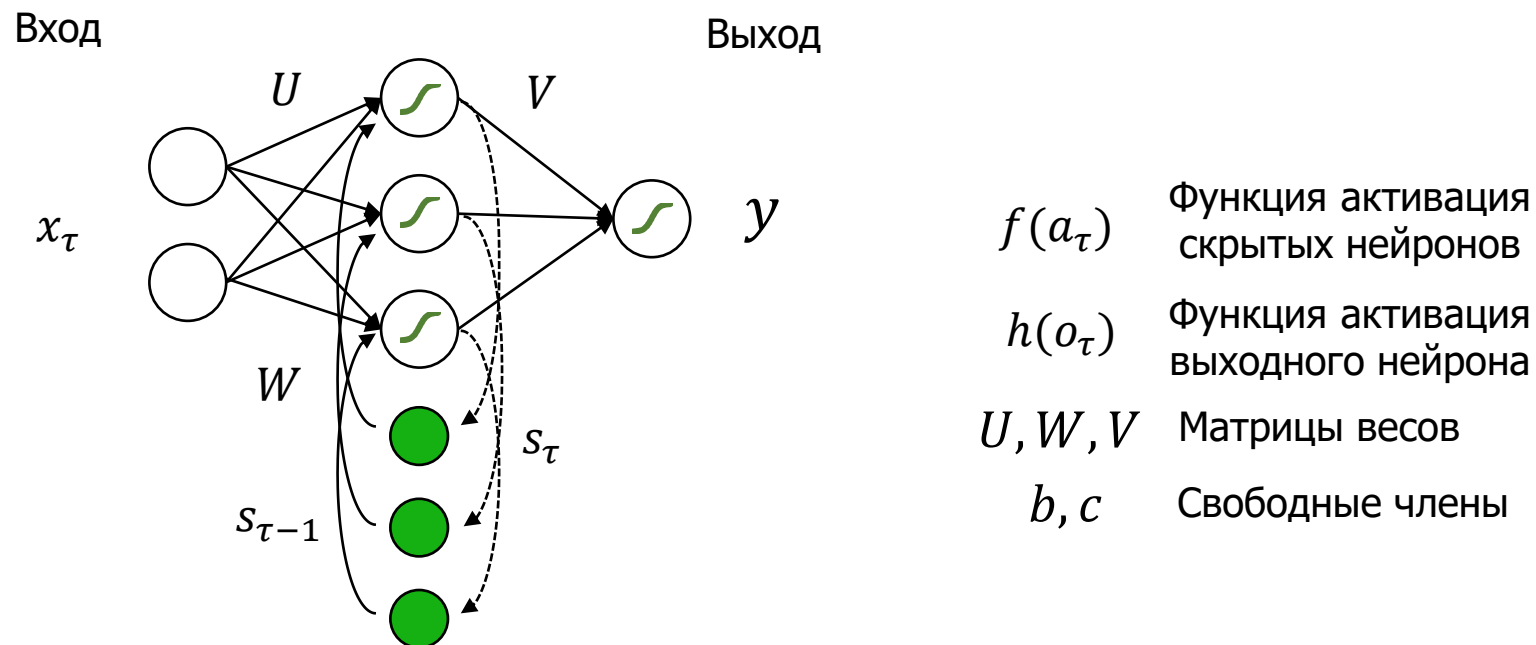


Синхронизированные
последовательности
входов и выходов



Последовательность входов и
последовательность выходов

Простая рекуррентная модель



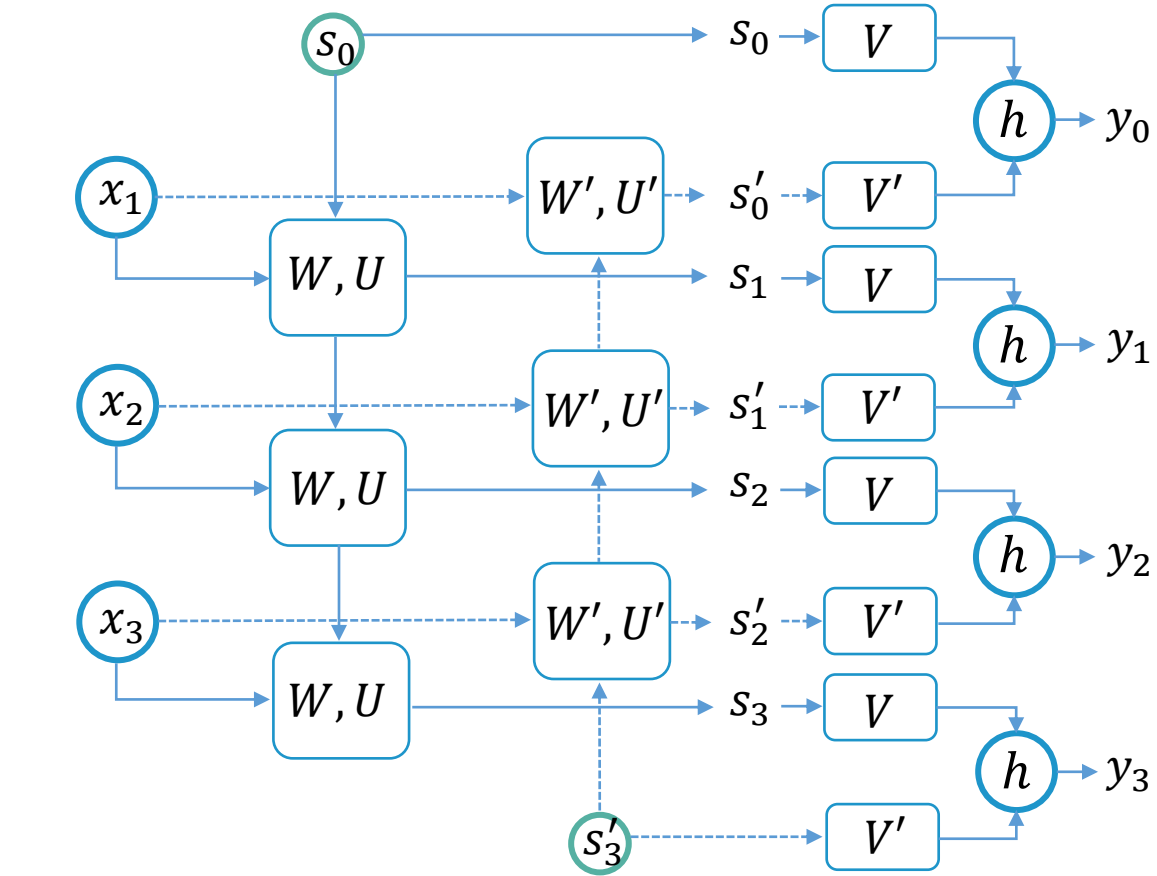
Активация скрытых нейронов

$$a_\tau = b + U \cdot x_\tau + W \cdot s_{\tau-1} \quad s_\tau = f(a_\tau)$$

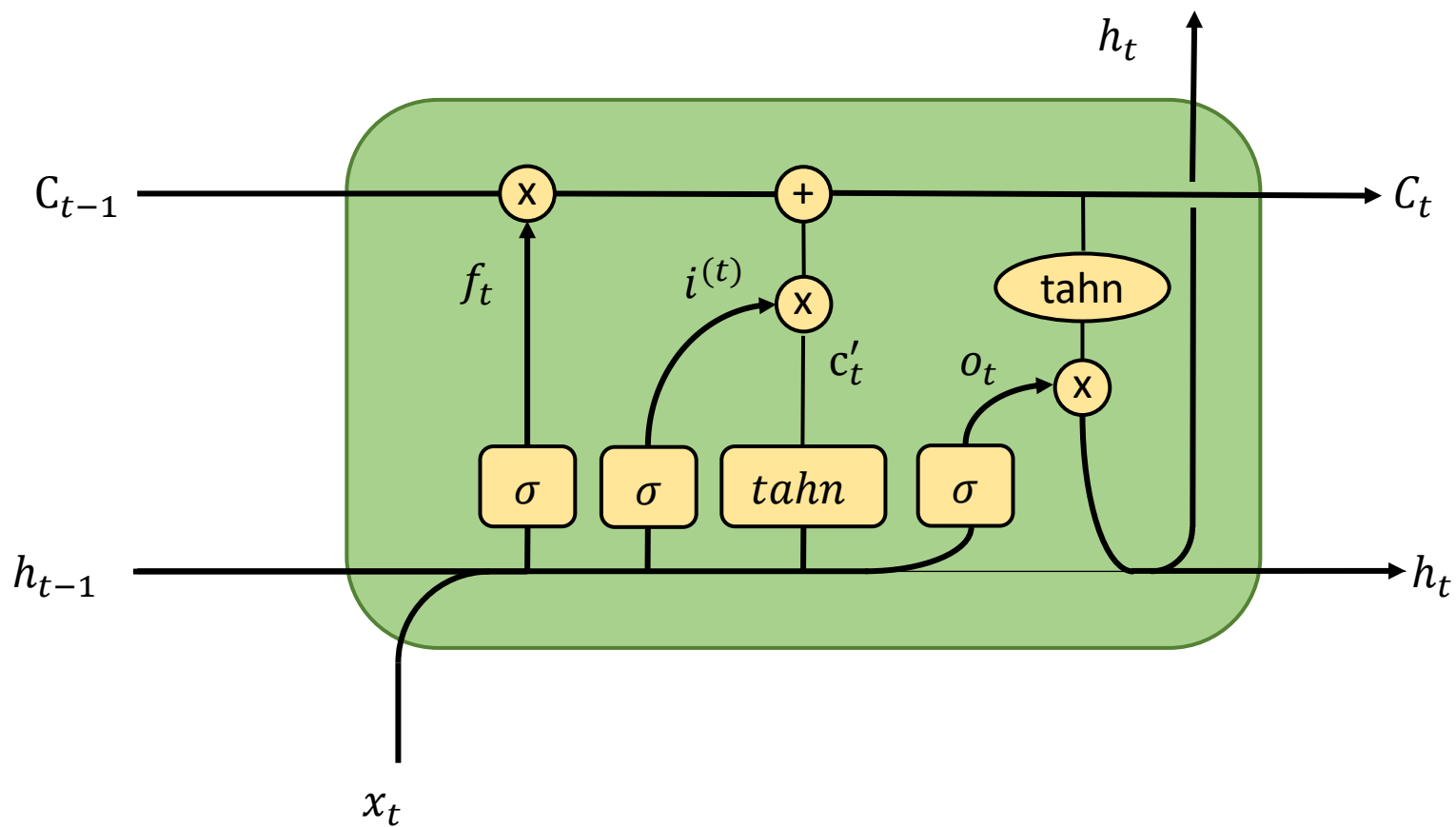
Активация выходного нейрона

$$o_\tau = c + V \cdot s_\tau \quad y_\tau = h(o_\tau)$$

Двунаправленная рекуррентная нейросеть



Долгая кратковременная память (LSTM)



LSTM: математика

Кандидат в новую ячейку памяти $c'_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_{c'})$

Входной гейт

$$i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + b_i)$$

Гейт забывания

$$f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + b_f)$$

Выходной гейт

$$o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + b_o)$$

Состояние ячейки

$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t$$

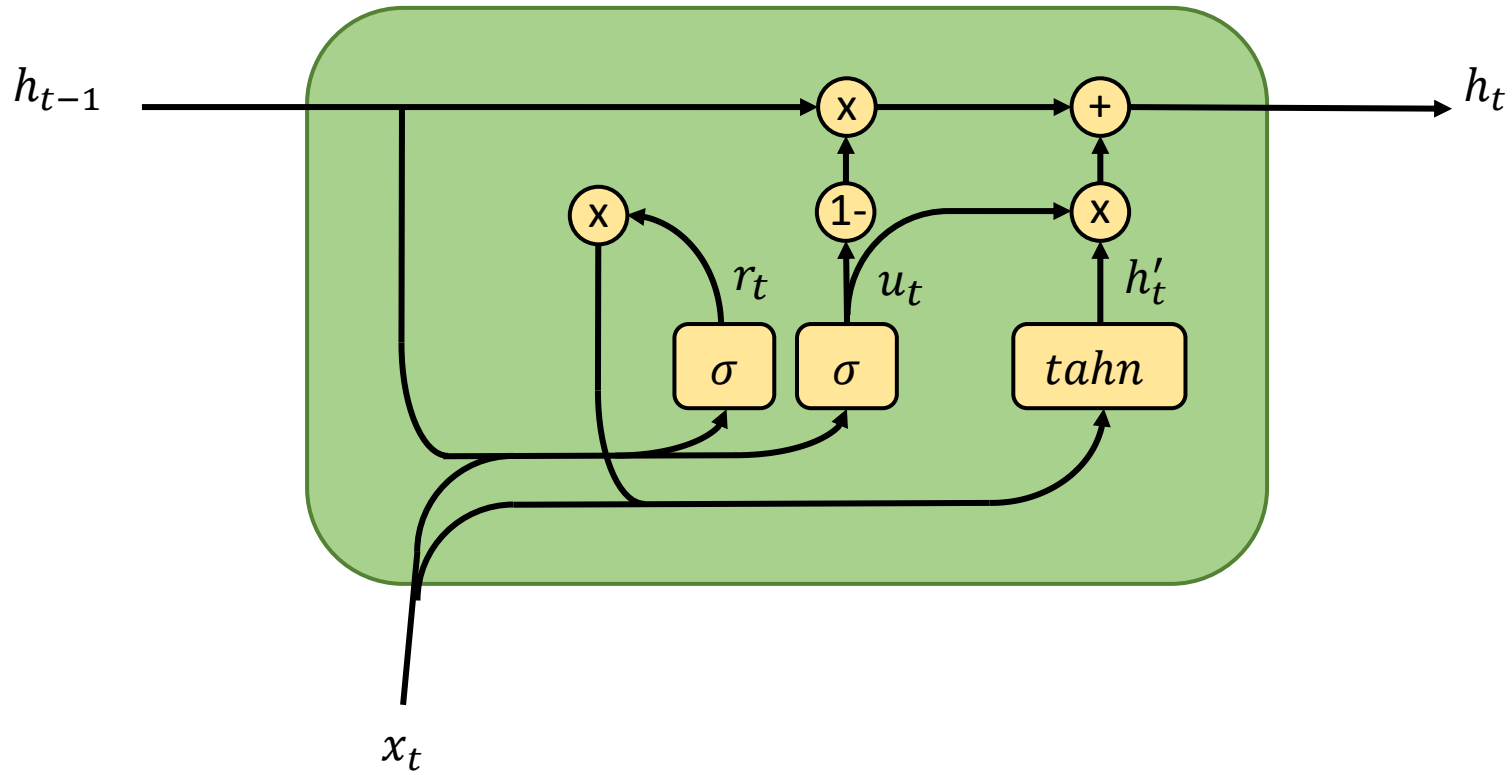
Выход блока

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

W - Матрица весов

$\odot$ - Поэлементное умножение

Gated Recurrent Unit (GRU)



GRU: математика

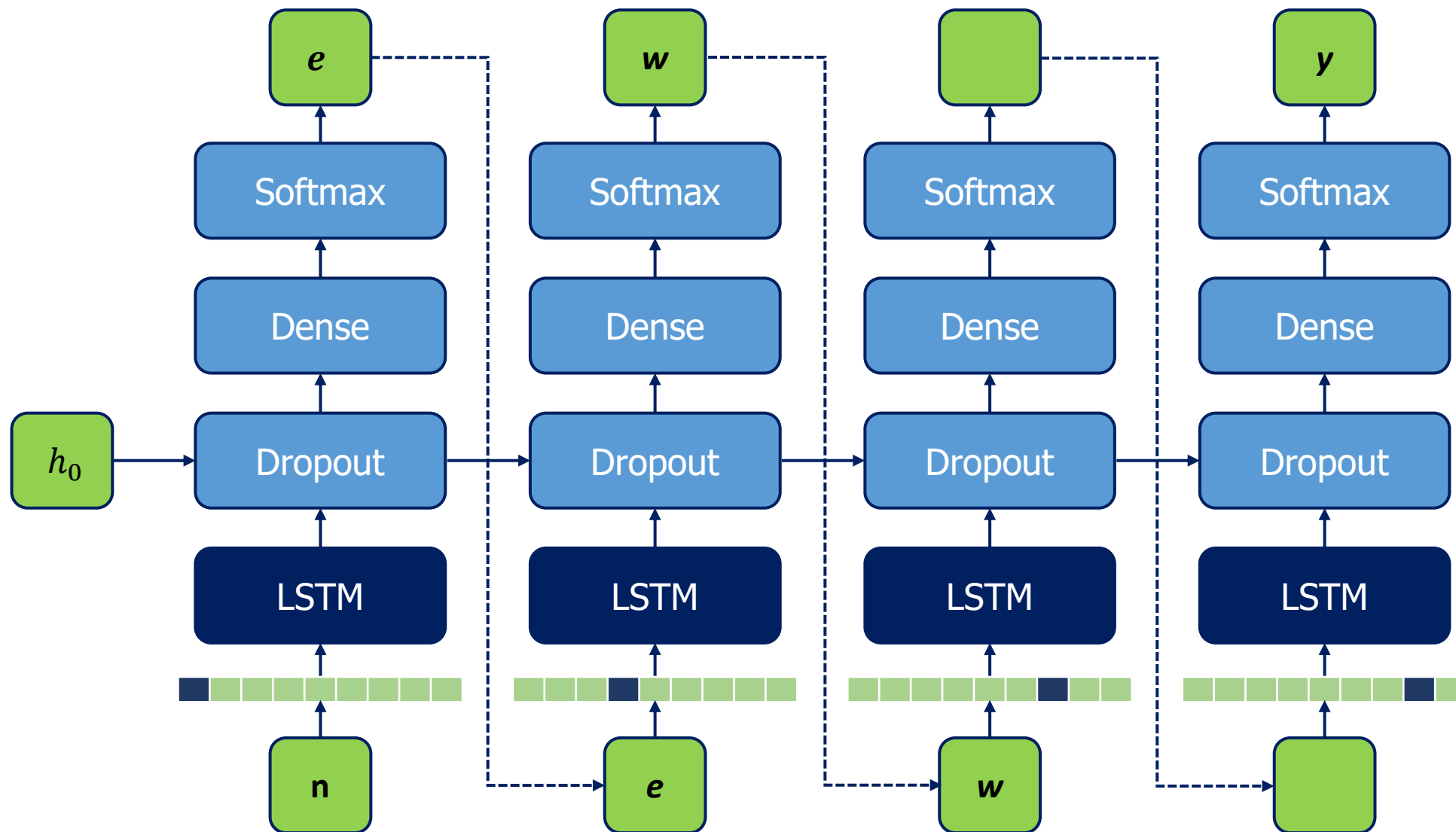
Гейт обновления $u_\tau = \sigma(W_{xu} \cdot x_\tau + W_{hu} \cdot h_{\tau-1} + b_u)$

Гейт перезагрузки $r_\tau = \sigma(W_{xr} \cdot x_\tau + W_{hr} \cdot h_{\tau-1} + b_r)$

Кандидат в выход $h'_\tau = \tanh(W_{xh'} \cdot x_\tau + W_{hh'} \cdot (r_\tau \odot h_{\tau-1}))$

Выход блока $h_\tau = (1 - u_\tau) \odot h'_\tau + u_\tau \odot h_{\tau-1}$

Пример архитектуры с одной ячейкой LSTM



Архитектура со стеком LSTM с остатками

