

Deep Learning for NLP

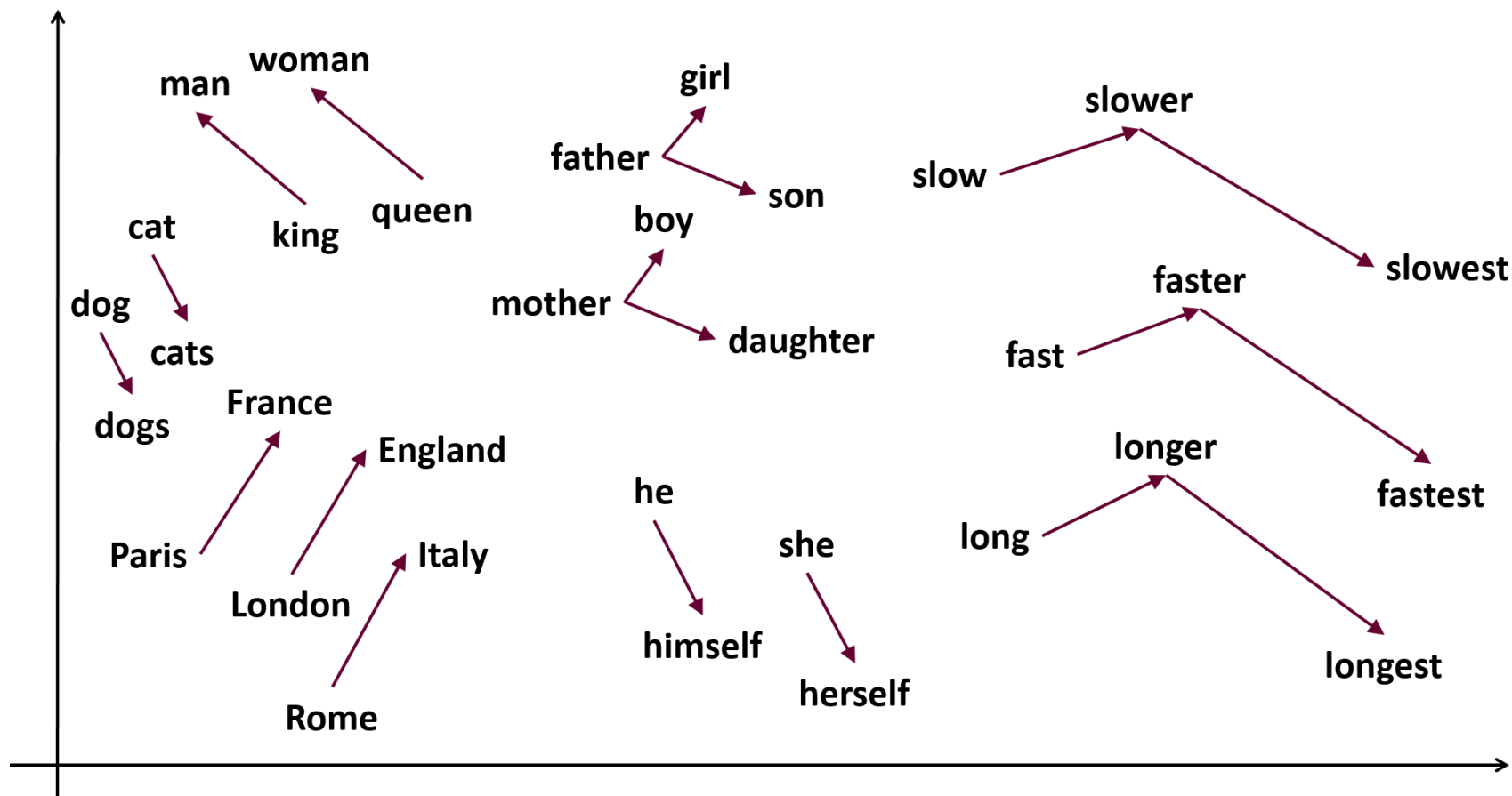
Введение

Даниил Анастасьев

dan.anastasyev@gmail.com

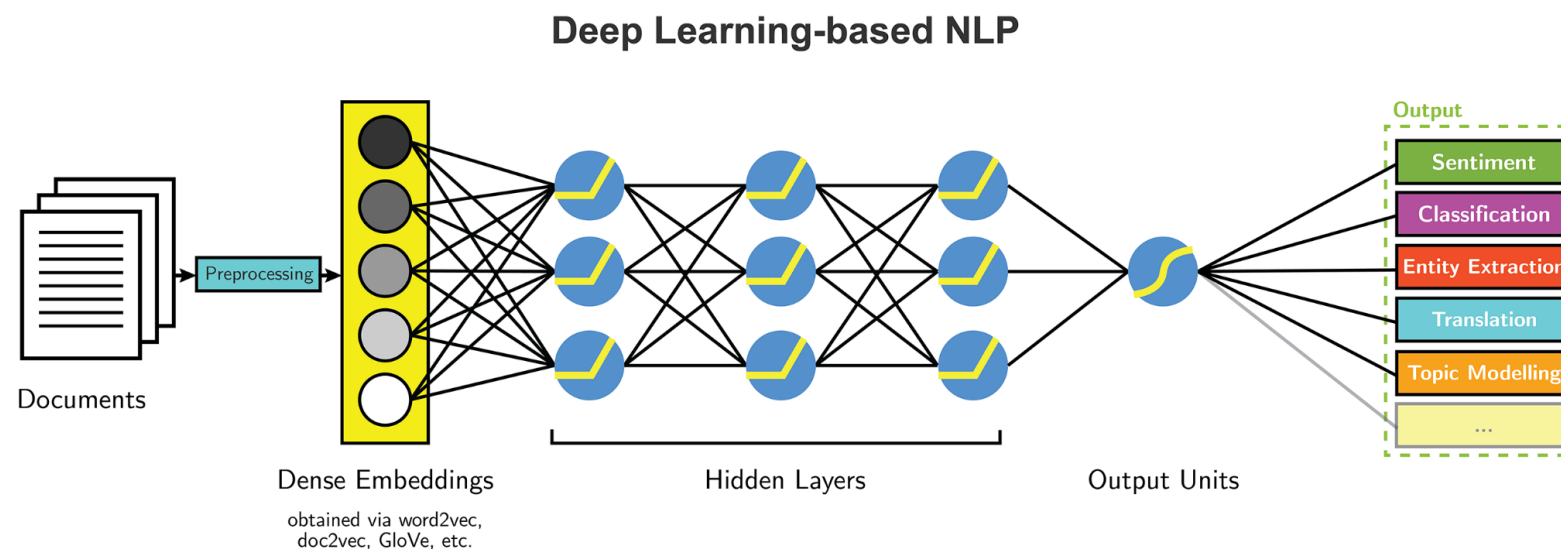
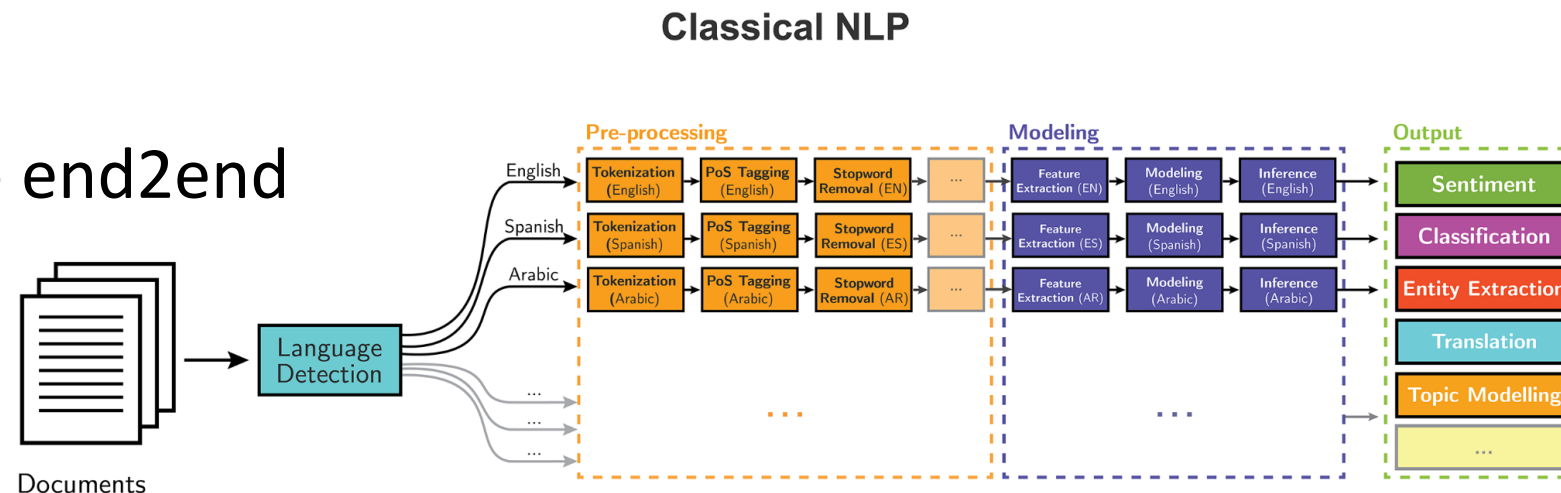
Почему DL?

- Потому что словные вектора



Почему DL?

- Потому что end2end



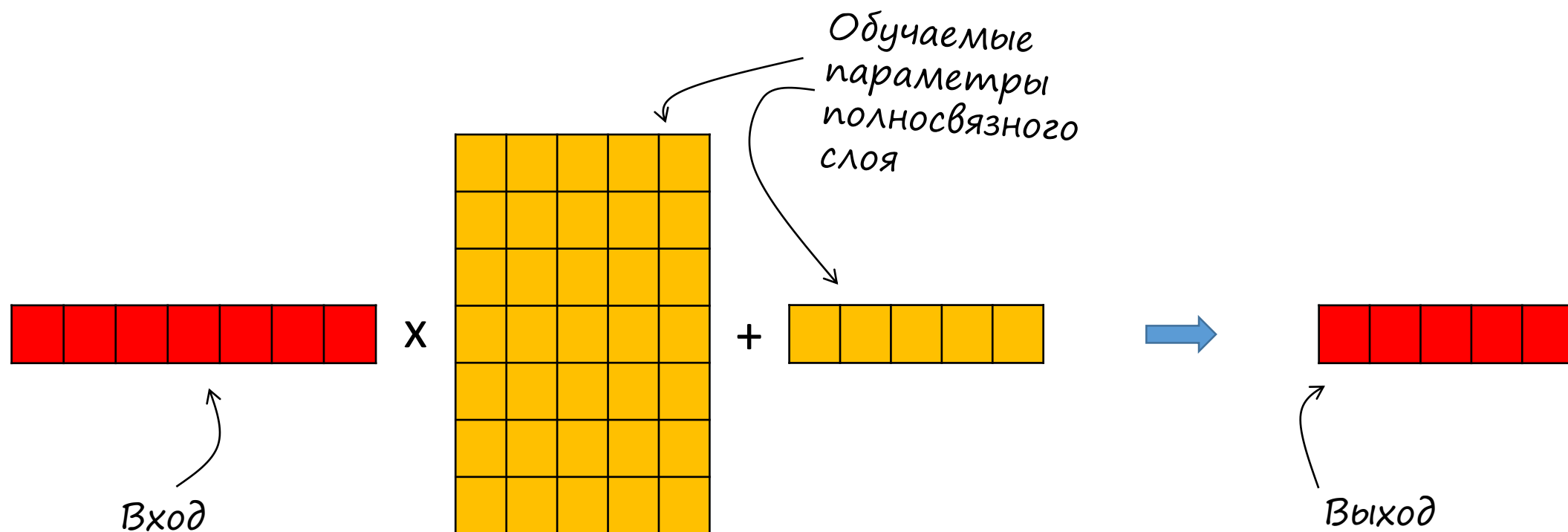
Элементы уважающих себя сетей

- Слои (и связи между ними)
- Функция потерь
- Оптимизатор

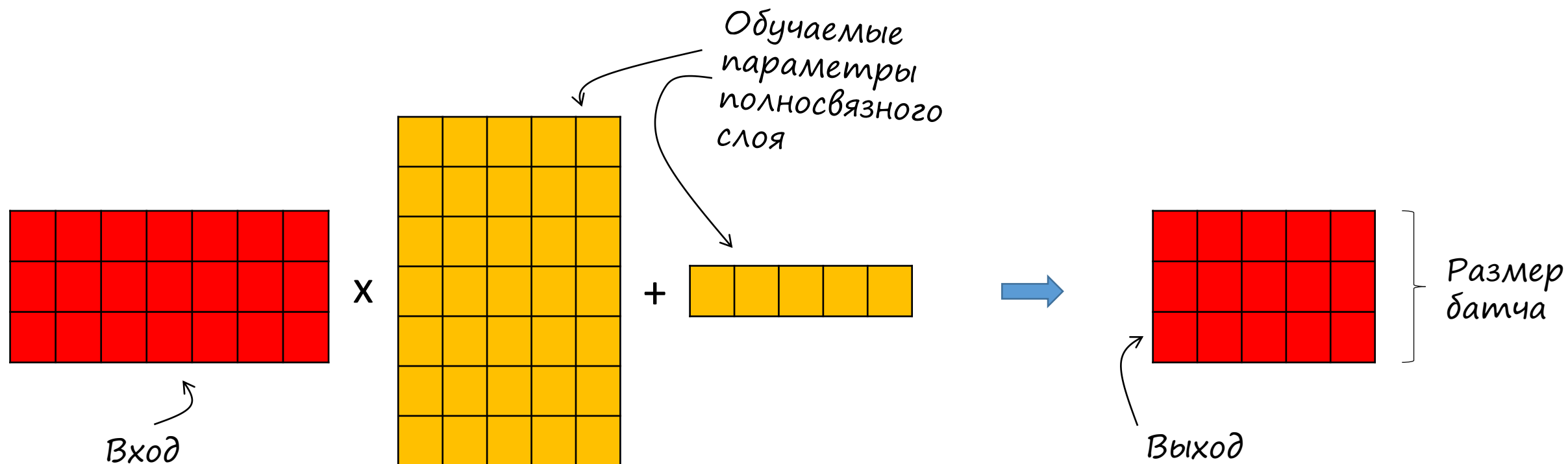
Основные слои

- Полносвязные
- Эмбединги
- Сверточные
- Рекуррентные
- + Активации
- + Слои регуляризации

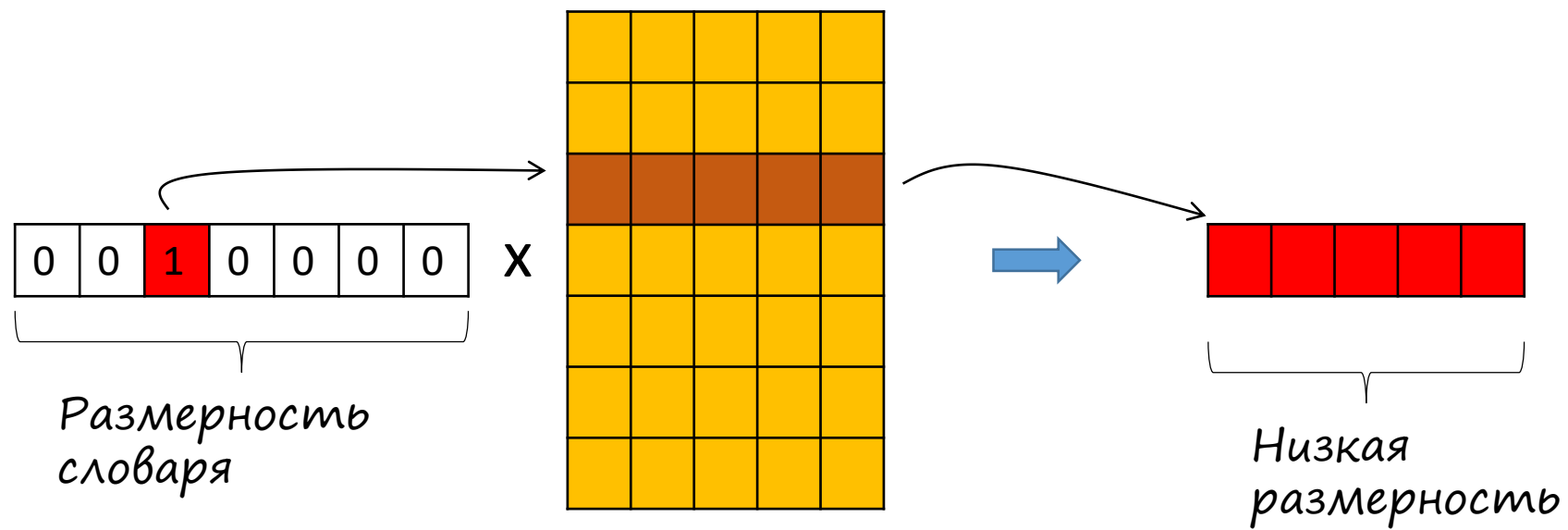
Полносвязные слои



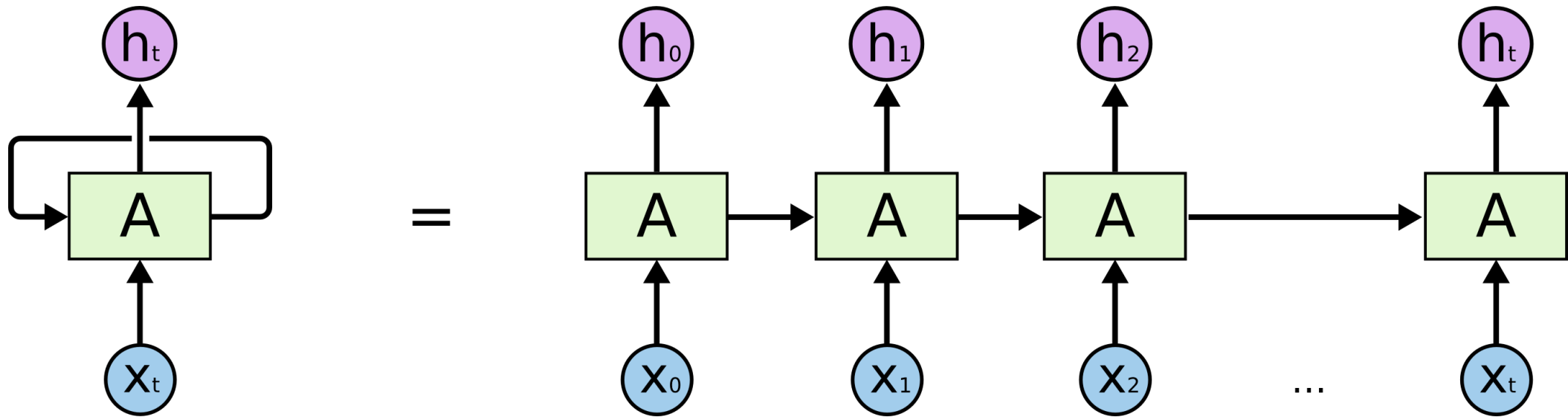
Полносвязные слои



Слой эмбедингов

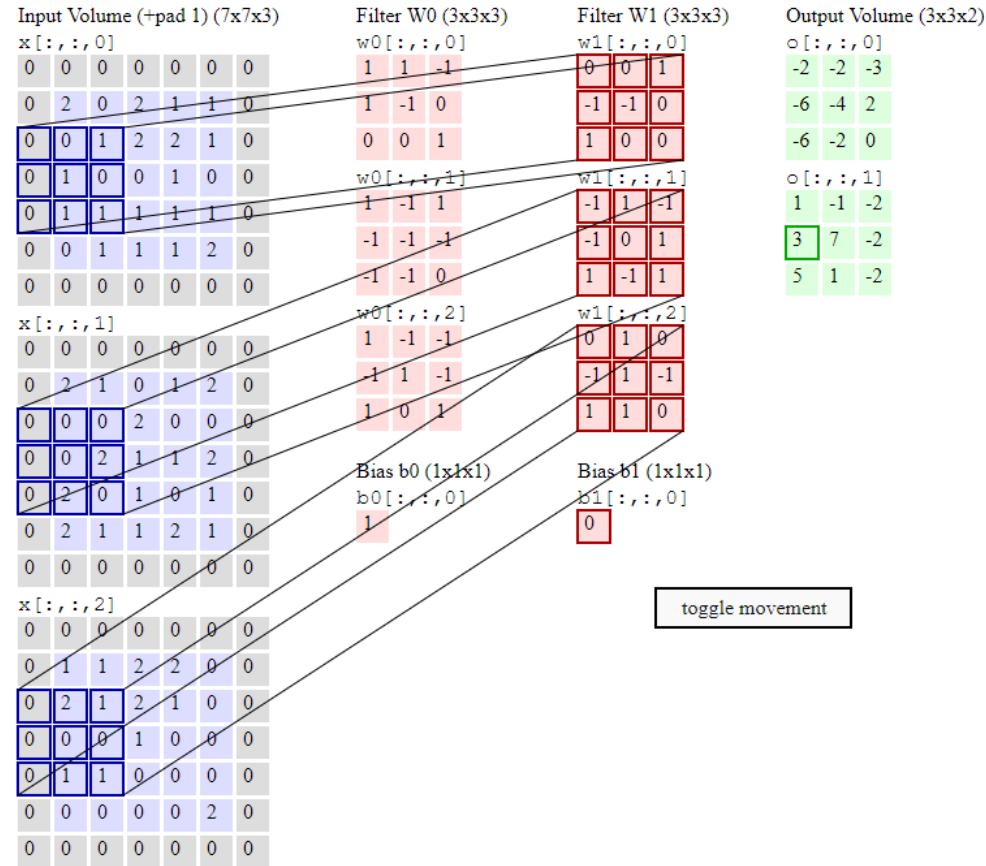


Рекуррентные слои



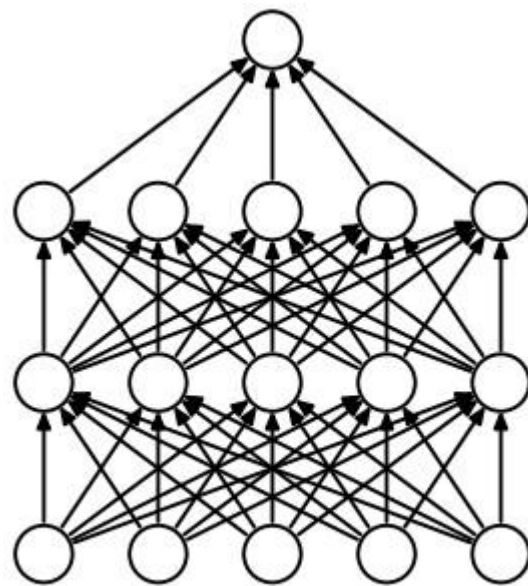
<http://colah.github.io/posts/2015-08-Understanding-LSTMs>

Сверточные сети

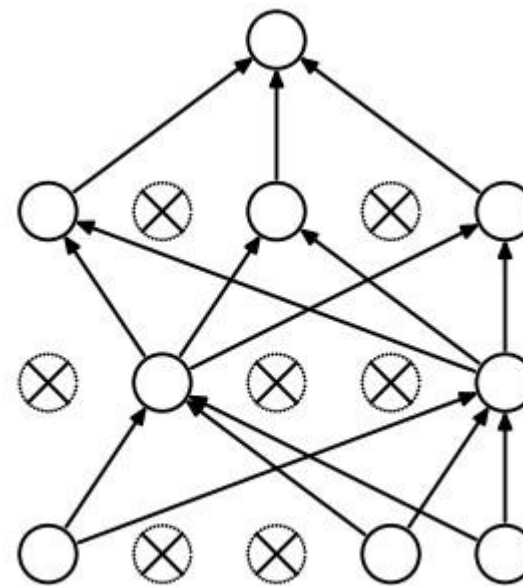


<http://cs231n.github.io/convolutional-networks>

Dropout

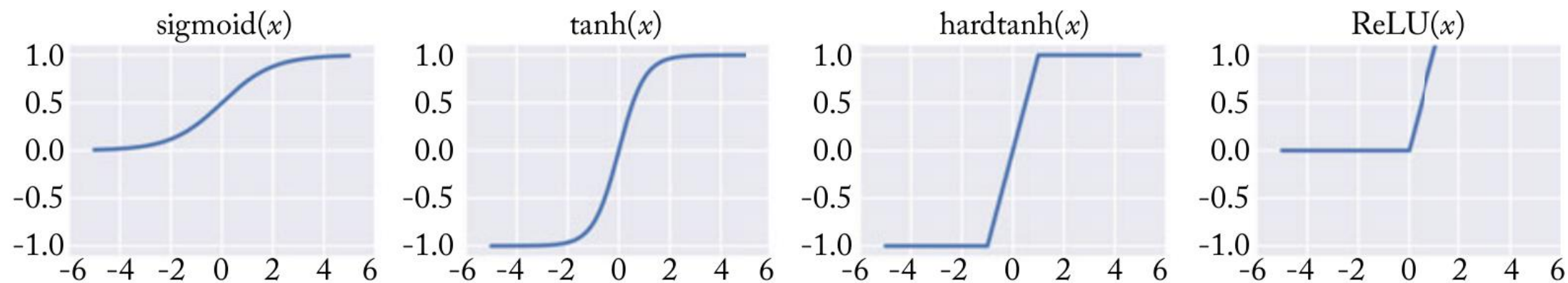


(a) Standard Neural Net



(b) After applying dropout.

Нелинейности



Функции потерь

Бинарная классификация:

$$L_{\text{logistic}}(\hat{y}, y) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y})$$

Многоклассовая классификация:

$$L_{\text{cross-entropy}}(\hat{y}, y) = - \sum_i y_{[i]} \log(\hat{y}_{[i]})$$

Оптимизаторы

SGD:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta)$$

В большинстве случаев используют адаптивные методы.
Чаще всего – **Adam**.

<http://runder.io/optimizing-gradient-descent/>