Нейронные сети

# Что такое нейронная сеть и как она работает

## Мотивация

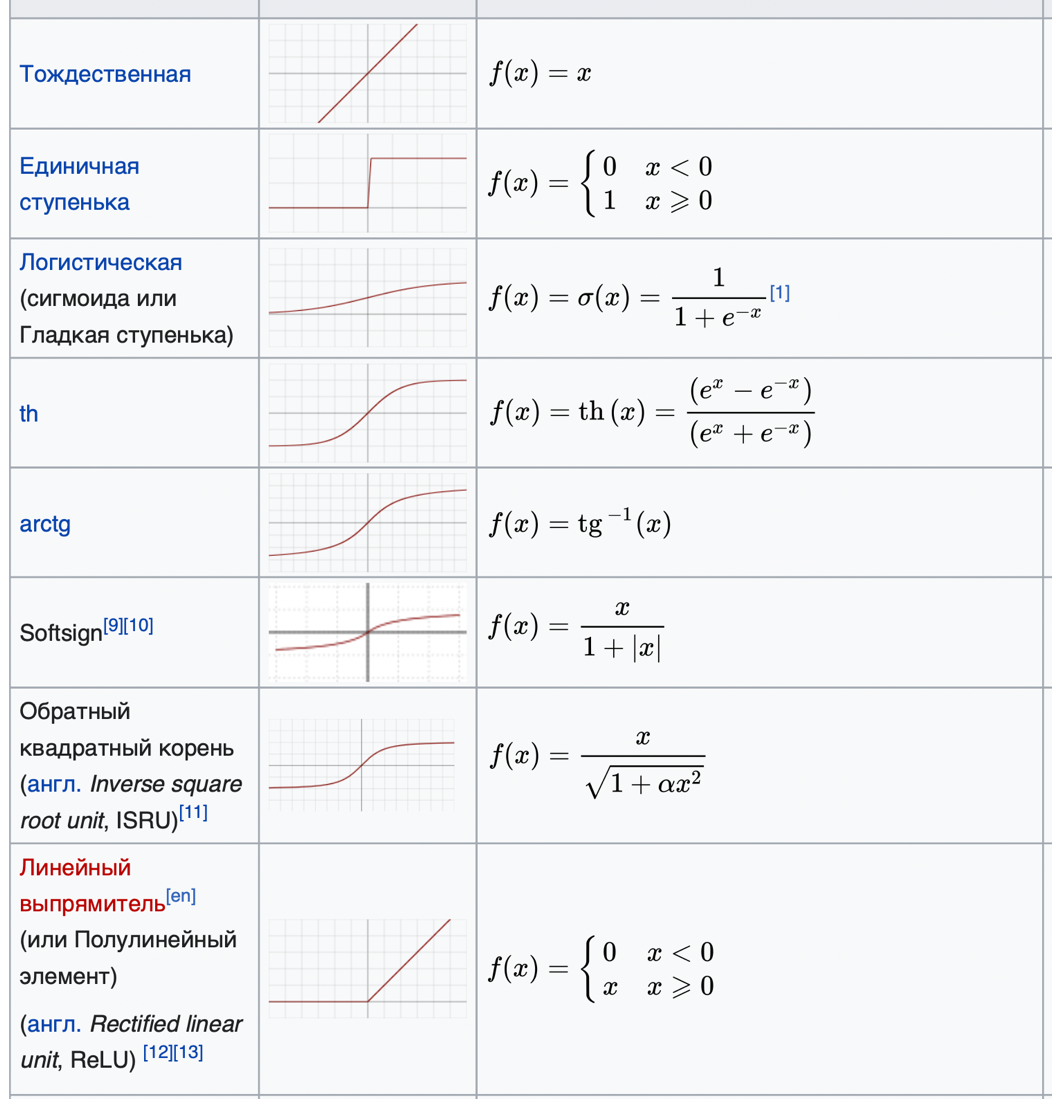
Человеческий мозг – удивительно сложный объект, способный решать очень большое количество сложных задач одновременно. Идея искусственных нейронных сетей пришла из верхнеуровневого понимания работы человеческого мозга. В нем около 8.6×1010 нейронов, но ещё больше (~1.5×1014) связей между ними.

Нейрон представляет из себя элемент, который̆ вычисляет выходной̆ сигнал (по определенному правилу) из совокупности входных сигналов. То есть основная последовательность действий одного нейрона такая:

* Прием сигналов от предыдущих элементов сети
* Комбинирование входных сигналов
* Вычисление выходного сигнала
* Передача выходного сигнала следующим элементам нейронной сети

## Модель

Рассмотрим простейшую математическую модель естественного нейрона. На рисунке ниже в центре размещен рассматриваемый нейрон, а слева изображены нейроны, которые посылают в него входной сигнал. Каждый из четырёх нейронов слева посылает сигнал, обозначаемый x1, x2, x3, x4. Каждый сигнал умножается на коэффициент w1, w2, w3, w4, характеризующий качество связи между двумя нейронами. 

Рассматриваемый нейрон собирает все взвешенные сигналы (wixi) использует их в качестве входа для функции активации (примеры ниже).

Фигура: примеры функций активации (данные из Wikipedia).

Выход функции активации и будет выходом нашего нейрона. Можно перейти от рассмотрения одного нейрона на выходе к набору нейронов и получить простейшую полносвязную нейросеть, состоящую из одного слоя.

## Основные понятия

* **Вход нейронной сети** – данные/ сигнал, которые нейросеть использует в качестве входного. Чаще всего на вход нейронной сети подают данные из обучающей выборки и хотят, чтобы выход нейронной сети был как можно ближе к меткам (таргетам) соответствующих данных. Это может быть картинка, аудиосигнал или вектор признаков.
* **Нейрон** – примитивный элемент нейросети, задача которого получить входные сигналы (x), просуммировать их с весами (w) и пропустить через функцию активации.
* **Веса нейронной сети** – параметры нейронной сети, определяющие её работу. Наряду с архитектурой (топологией) нейронной сети.
* **Архитектура (топология) нейронной сети** - способ соединения нейронов в слои и слоев между собой. Обычно, чем больше слоев и нейронов, тем более сложные зависимости может выучить нейросеть, однако, тем сложнее её обучать.
* **Обучение нейронной сети** – подбор весов нейросети таким образом, чтобы минимизировать функцию потерь на обучающей выборке. Это самый тяжелый с точки зрения вычислительных ресурсов и временных затрат этап. Как правило, нейросети обучаются с помощью алгоритмов типа стохастического градиентного спуска. Во время обучения нейросети показывают группы из обучающей выборки, считают значение функции потерь и градиенты функции потерь по весам нейронной сети (с помощью алгоритма обратного распространения ошибки) и обновляют эти веса, чтобы уменьшить значение функции потерь.
* **Батч** (мини-батч) – набор объектов из обучающей выборки, на которые нейросеть «смотрит» за одну итерацию.
* **Эпоха** – количество итераций обновлений весов, после которых нейросеть «увидела» все данные обучающей выборки один раз.

# Типы нейронных сетей и решаемые задачи

|  |  |
| --- | --- |
| Тип данных/ решаемая задача | Модели |
| Классификация изображений | LeNet, AlexNet, ResNet, VGG, GoogleNet, ViT |
| Генерация изображений | StyleGAN, Diffusion models |
| Моделирование естественного языка | GPT-3, Megatron-LM |
| Распознавание речи | Conformer, BERT, LSTM |

## Свёрточные нейронные сети

Ключевые слова: AlexNet, LeNet, GoogleNet, ResNet

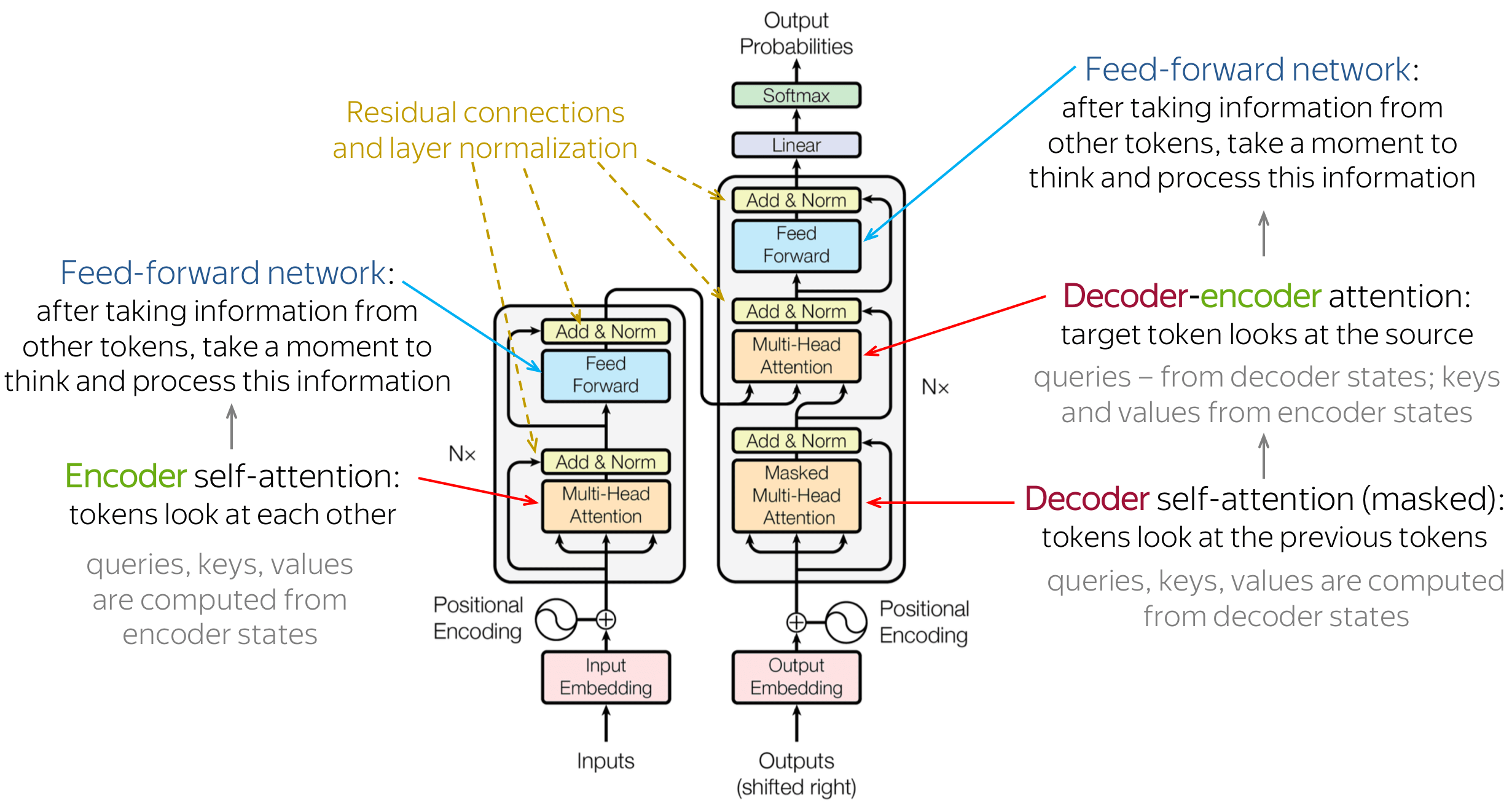
## Рекуррентные нейронные сети

Ключевые слова: LSTM, GRU, Transformers (частично, о них - далее)

## Генеративные нейронные сети

Ключевые слова: GAN, StyleGAN, CyclicGAN, DALL-E, Midjourney

## Архитектуры Трансформерного типа



Ключевые слова: GPT (GPT, GPT-2, GPT-3), OPT, YaLM, T5

# Данные

## Инструменты для получения наборов данных

* Yandex Toloka
* Amazon Turk

## Использование собственных данных

* Transfer learning
* Promt engineering

# На чем обучают нейронные сети

## Видеокарты

Для обучения нейронных сетей обычно используют GPU, потому что они состоят из большого количества вычислительных ядер, что позволяет использовать параллелизм вычислений.

## Собственные кластеры

* NVIDIA DGX

## Арендные мощности

* Sbercloud
* Amazon AWS
* Vast.ai
* Google cloud

# Полезные ссылки

* [Paperswithcode](https://paperswithcode.com/)
* [Thisxdoesnotexist](https://thisxdoesnotexist.com/)