Введение в нейронные сети

Владимир Литвиненко

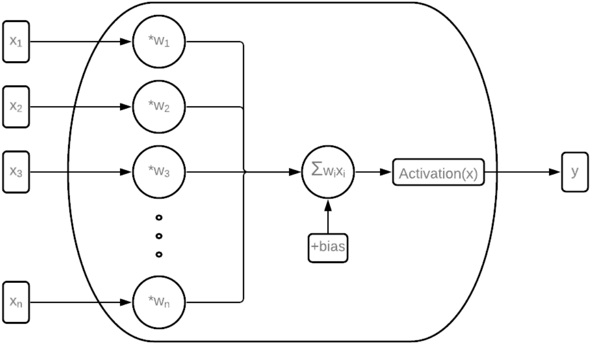
Московский физико-технический институт

23 ноября 2020 г.

План

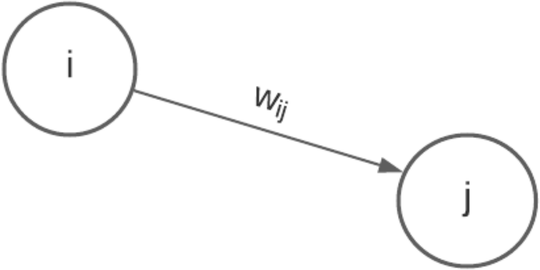
1. Связи нейронов
2. Функции активации
3. Слои

Нейрон



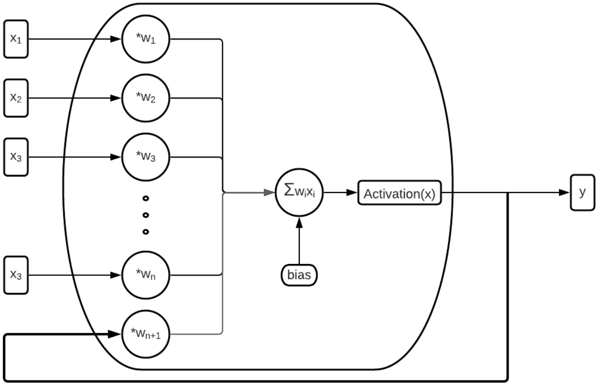
$$y = Activation(\sum_{j=1}^{n} w_j x_j + bias)$$

Веса и связи



Обратная связь

Рекуррентные нейронные сети (RNN) построены на обратной связи

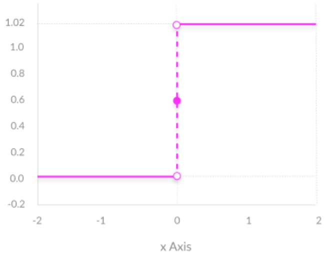


Работают с потоком данных

Threshold function

$$f(x) = \begin{cases} 1 & \text{if } x \geq \textit{threshold} \\ 0 & \text{if } x < \textit{threshold} \end{cases}$$

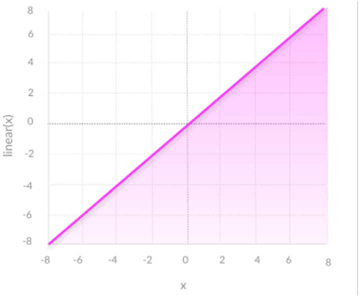
- Подходит только для классификации
- Имеет разрыв. Нехорошо для обучения



Linear function

$$f(x) = x$$

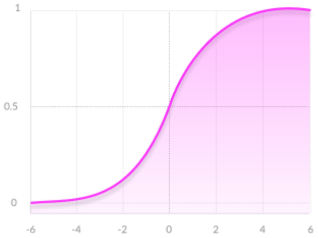
- Подходит для классификации и регрессии
- Легко считается производная
- Линейная :)
- Производная не зависит от сигнала



Sigmoid function

$$f(x) = \frac{1}{1+e^{-x}}$$

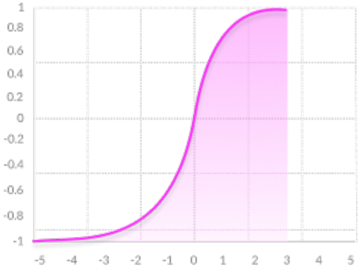
- Подходит для классификации и регрессии
- Нелинейная
- Вычисление производной - дорогой процесс
- Производная зависит от сигнала



Tanh function

$$f(x) = \frac{2}{1+e^{-}2x} - 1$$

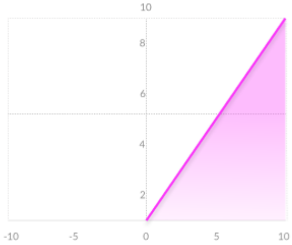
- Аналогична сигмоиде, но с другими границами



ReLu

$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

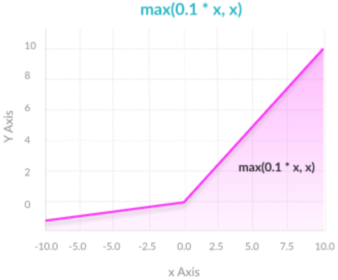
- Универсальна
- Легко считается производная
- Нелинейна
- При отрицательном сигнале нейрон не обучается



Leaky ReLu

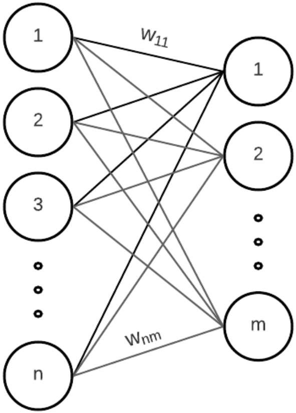
$$f(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0.1x & \text{if } x < 0 \end{cases}$$

- Аналогична ReLu
- Отсутствие проблемы умирающих нейронов

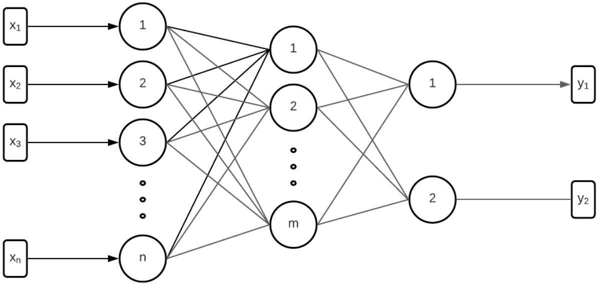


Fully connected network (FCN)

- Один нейрон линейно разделяет область
- Двухслойная сеть ограничивает выпуклую многогранную область
- Трехслойная сеть ограничивает многогранную область необязательно выпуклую, необязательно связную



Многоканальный выход



Теорема Цыбенко (1989)

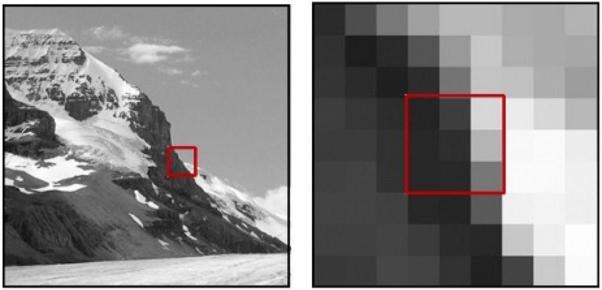
Если $\sigma(\vec{x})$ - непрерывная сигмоида, то для любой непрерывной на $[0,1]^n$ функции $f(\vec{x})$ существуют такие значения параметров $\vec{w_j} \in \mathbb{R}^n, bias_j \in \mathbb{R}$, что двухслойная сеть

$$a(x) = \sum_{j=1}^{n} \alpha_j \sigma(\vec{w_j}^T \vec{x} + bias_j)$$

равномерно приближает $f(\vec{x})$ с любой точностью ε или

$$|a(\vec{x}) - f(\vec{x})| < \varepsilon, \text{ для всех } \vec{x} \in [0,1]^n$$

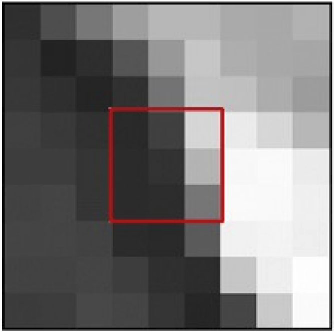
Проблема анализа изображений



43	102	169
35	58	191
38	44	155

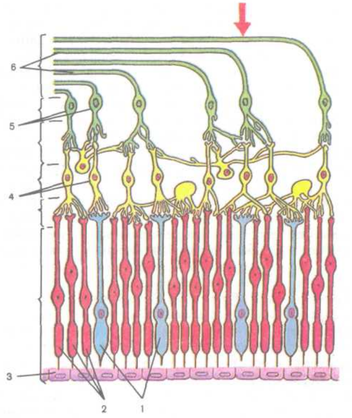
Проблема анализа изображений

- Очень много данных. Пример: изображение 10x10 - это 100 входов
- Данные неинформативные
- Вывод: очень трудно подобрать веса

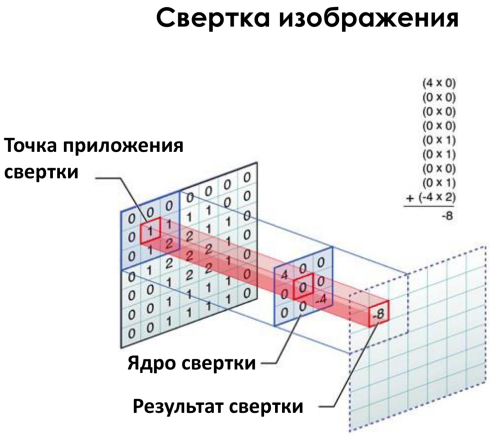


Сверточный слой. История

- Идея сетчатки
- Нейрон подходит к области сетчатки, а не к каждому сенсору отдельно
- Агрегирует информацию и выделяет более информативные источники



Сверточный слой. Наглядно



Сверточный слой. Формула

$$v_{k,m} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}x_{i+k,j+m} + bias,$$

где k,m = 0, 1, 2, ... - индексы результата; n - размер ядра свертки; i, j = 0, 1, ..., n - индексы ядра свертки