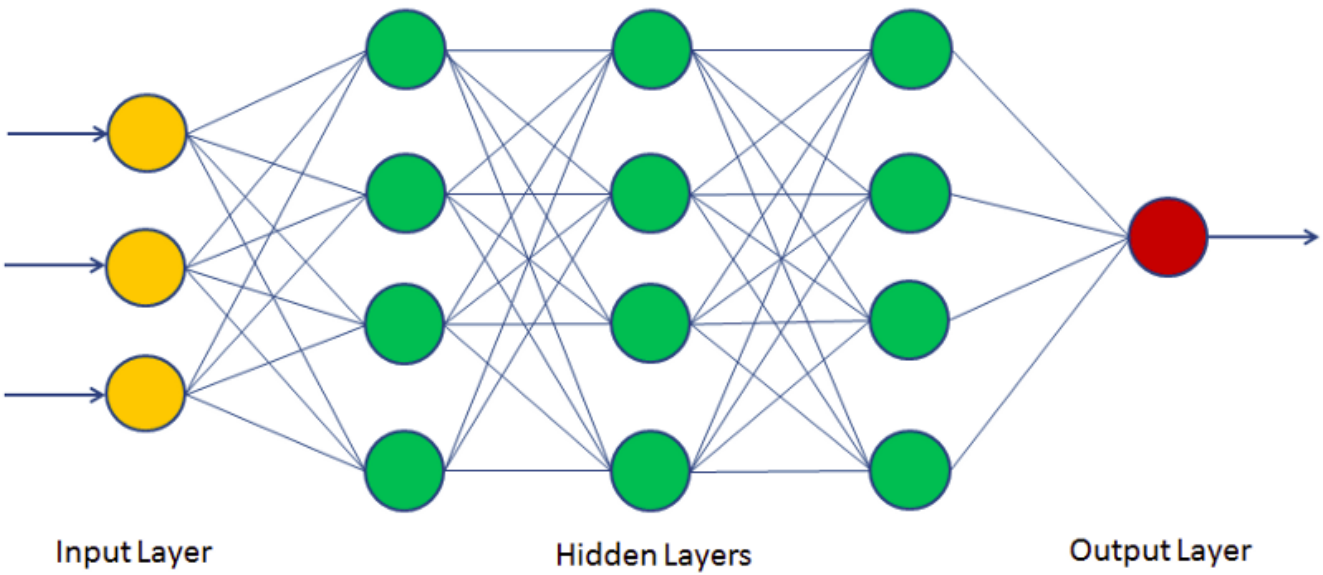
Многослойный персептрон

Виктор Китов
victorkitov.github.io



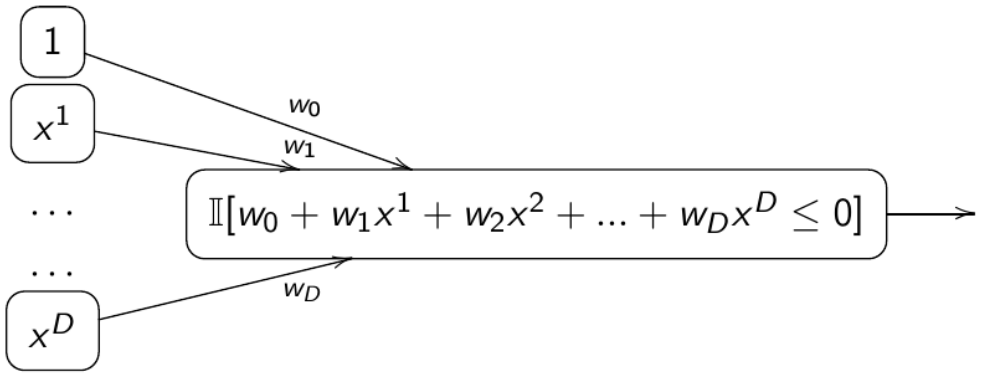
Содержание

История

- Нейронные сети появились как попытка моделировать работу человеческого мозга.

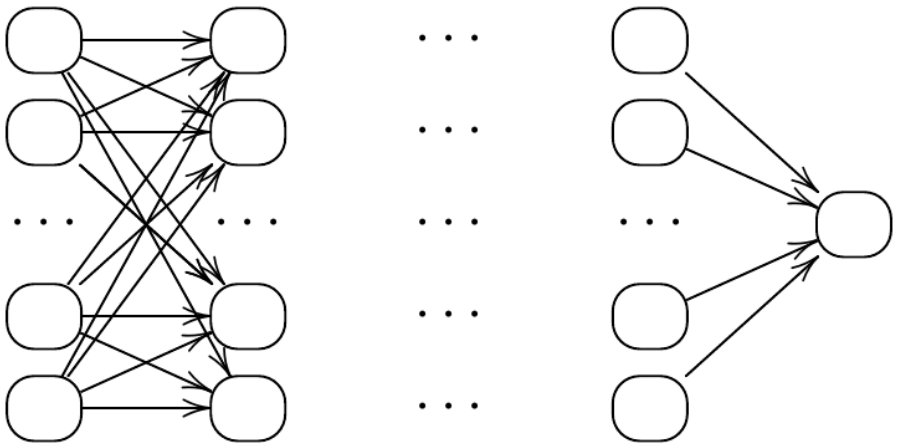
- Человеческий мозг состоит из взаимосвязанных нейронов.
  - порядка 86 миллиардов нейронов
- нейроны связаны аксонами - вытянутыми отростками нервных клеток
- взаимодействие нейронов осуществляется электро-химическими сигналами по аксонам

Простая модель нейрона



- Несколько входов посылают сигналы, которые домножаются на вес связи
- Нейрон принимает суммарный сигнал
- Нейрон активируется в полупространстве $w_0 + w_1 x^1 + w_2 x^2 + ... + w_D x^D \leq 0$.
- w_0 отвечает за смещение

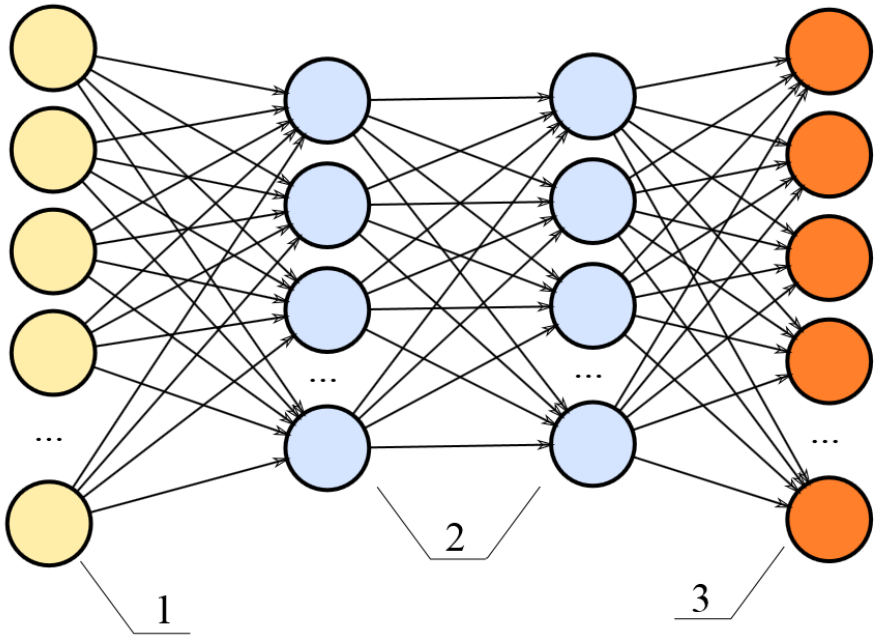
Архитектура многослойного персептрона



многослойный персептрон - ациклический направленный граф

- Несколько слоев, связи между соседними слоями - каждый с каждым.
- Каждый нейрон имеет свои собственные связи.

Слои



- Слои многослойного персептрона:
  - 1-входной слой (не учитывается в полном количестве слоев сети)
  - 2-скрытые слои
  - 3-выходной слой

Многослойный персептрон и ансамбли

- В стэкинге фиксируются базовые модели при настройке агрегирующей ф-ции.
- В бустинге фиксируются предыдущие базовые модели.
- В многослойном персептроне ранние и поздние нейроны настраиваются одновременно.
  - более сильное переобучение

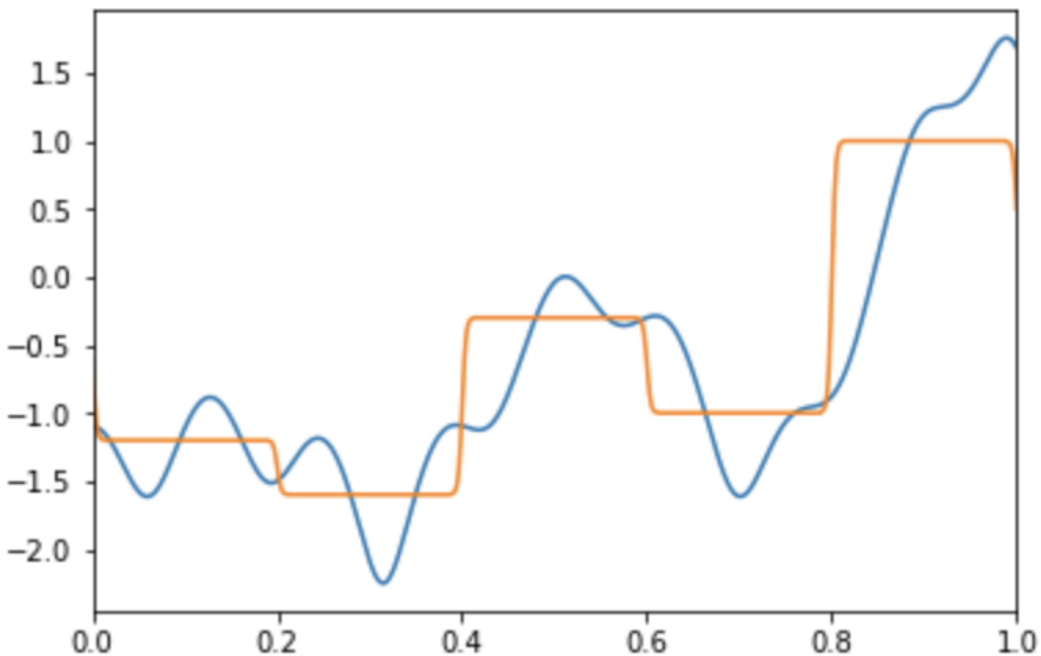
Содержание

Одномерная регрессия

- 1-мерная регрессия:

$$f(x) = \sum_i f(b_i)\mathbb{I}[x \in (b_i, b_{i+1}]] = \sum_i f(b_i)\left(\mathbb{I}[x \leq b_{i+1}] - \mathbb{I}[x \leq b_i]\right)$$

$$= \sum_i f(b_i)\mathbb{I}[x \leq b_{i+1}] - \sum_i f(b_i)\mathbb{I}[x \leq b_i]$$ 2-х слойный персептрон



Многомерная регрессия

- AND/OR функции для $x_1, x_2 \in \{0, 1\}$ можно сделать 1 слойным персептроном:[1]:

$$\text{AND function} \mathbb{I}[x_1 + x_2 \geq 2] = \mathbb{I}[-x_1 - x_2 \leq -2]$$
$$\text{OR function} \mathbb{I}[x_1 + x_2 \geq 1] = \mathbb{I}[-x_1 - x_2 \leq -1]$$

- **D-мерная регрессия:**
  - один слой приближает линейную ф-цию
    - 2-х слойный персептрон моделирует индикаторную ф-цию на произвольном выпуклом многоугольнике (через AND)
    - 3-х слойный персептрон приближает произвольную непрерывную функцию (Липшицеву)
      (как взвешенную сумму индикаторов выпуклых многоугольников)
- Таким образом, 3-х слоев достаточно для приближения всех регулярных зависимостей.

[1] How to make XOR (exclusive OR) function?

Классификация

- **Классификация:**
  - один слой выделяет полупространства
  - 2-х слойный персептрон моделирует индикаторную ф-цию на произвольном выпуклом многоугольнике (через AND)
    - приближает произвольное выпуклое множество
  - 3-х слойный персептрон выделяет произвольный многоугольник (через OR) как объединение выпуклых многоугольников
- Таким образом, 3-х слоев достаточно для приближения любого множества.

Выбор числа слоёв

- Зачем использовать больше 3-х слоёв?
- 3-х слойные сети способны приближать любые регулярные зависимости, но может потребоваться слишком много нейронов - переобучение.
- Более глубокие слои могут переиспользовать ранние нейроны.
  - нужно меньше нейронов, меньше связей, меньше переобучение

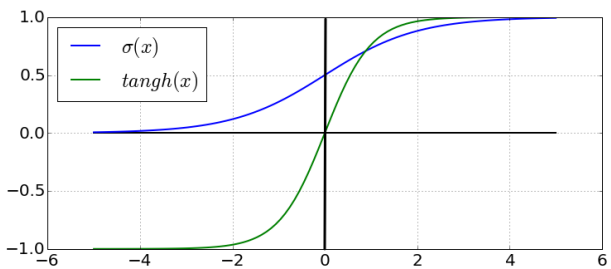
Содержание

Непрерывные активации

- $\mathbb{I}[w^T x - w_0 \leq 0]$ - кусочно-постоянная, производная=0, не можем оптимизировать веса.
- Заменим $\mathbb{I}[w^T x - w_0 \leq 0]$ непрерывной функцией активации $\phi(w^T x - w_0)$.

Основные функции активации

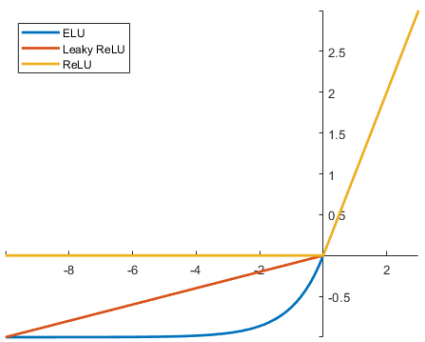
- сигмоида: $\sigma(x) = \frac{1}{1+e^{-x}}$
  - 1 нейрон с сигмоидой моделирует логистическую регрессию
- гиперболический тангенс: $\text{tangh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = 2\sigma(2x) - 1$
  - преимущество: если $\mathbb{E}x = 0$, то $\mathbb{E}\,\text{tangh}(x) = 0$.



- Проблема: $\phi'(x) \approx 0$ вне интервала (-3,3).

Основные функции активации

- Rectified linear unit (ReLU): $\phi(x) = \max(0, x)$
  - аналог с гладкой производной - SoftPlus: $\phi(x) = \ln(1 + e^x)$
- Leaky ReLU: $\phi(x) = \begin{cases} x, & x \geq 0 \\ 0.01x, & x < 0 \end{cases}$
- Parametric ReLU (α оценивается): $\phi(x|\alpha) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$
- Exponential LU (α задано): $\phi(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & x < 0 \end{cases}$



Содержание

Регрессия

- Регрессия: $\phi(I) = I$
- Скалярная регрессия $y \in \mathbb{R}$:

$$MSE(x, y) = \frac{1}{N} \sum_{n=1}^{N} (\widehat{y}(\mathbf{x}_n) - y_n)^2$$

$$MAE(x, y) = \frac{1}{N} \sum_{n=1}^{N} |\widehat{y}(\mathbf{x}_n) - y_n|$$

- Векторная регрессия $\mathbf{y} \in \mathbb{R}^K$:

$$MSE(x, y) = \frac{1}{N} \sum_{n=1}^{N} \|\widehat{\mathbf{y}}(\mathbf{x}_n) - \mathbf{y}_n\|_2^2$$

Классификация, вероятности классов

- **Бинарная классификация**: $y \in \{0, 1\}$

$$p(y = +1|x) = \frac{1}{1 + e^{-I}}$$

$$\mathcal{L}(x, y) = -\ln p(y|x) = -\ln p(y = 1|x)^{\mathbb{I}[y=1]}[1-p(y = 1|x)]^{\mathbb{I}[y\neq 1]}$$

Классификация, вероятности классов

- **Бинарная классификация**: $y \in \{0, 1\}$

$$p(y = +1|x) = \frac{1}{1 + e^{-I}}$$

$$\mathcal{L}(x, y) = -\ln p(y|x) = -\ln p(y = 1|x)^{\mathbb{I}[y=1]}[1-p(y = 1|x)]^{\mathbb{I}[y\neq 1]}$$

- **Многоклассовая классификация**: $y \in 1, 2, ...C$

$$\{SoftMax(I_1, ...I_C)\}_j = p(y = j|x) = \frac{e^{I_j}}{\sum_{k=1}^{C} e^{I_k}},\ j = 1, 2, ...C$$

$$\mathcal{L}(x, y) = -\ln p(y|x) = -\ln \prod_{c=1}^{C} p(y = c|x)^{\mathbb{I}[y=c]}$$

Классификация, рейтинги классов

- **Бинарная классификация**: $y \in \{-1, 1\}$

$$g(x) = \text{отн. предпочтительность положит. класса}$$

$$hinge(x, y) = [\alpha - yg(x)]_+$$

Классификация, рейтинги классов

- **Бинарная классификация**: $y \in \{-1, 1\}$

$$g(x) = \text{отн. предпочтительность положит. класса}$$

$$hinge(x, y) = [\alpha - yg(x)]_+$$

- **Многоклассовая классификация**: $y \in 1, 2, ...C$:

$$\{g_1(x), ...g_C(x)\} \text{ - рейтинги классов } 1, ...C$$

$$hinge_1(x, y) = \left[\max_{c \neq y} g_c(x) + \alpha - g_y(x)\right]_+$$

$$hinge_2(x, y) = \sum_{c \neq y} [g_c(x) + \alpha - g_y(x)]_+$$

- $\alpha > 0$ - подбираемый гиперпараметр (baseline: 1)

Заключение

- Нейросети - универсальный аппроксиматор: может моделировать сложные нелинейные зависимости.
- ReLU, LeakyReLU - рекомендуемые функции нелинейности.
- Функции потерь:
  - регрессия: $(\widehat{y}-y)^2, |\widehat{y}-y|$
  - классификация: кросс-энтропийные потери, hinge.