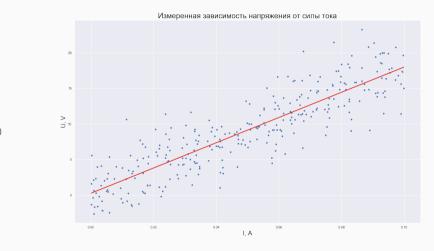
Лекция 4: Градиентный спуск и искусственные нейроны

20 марта 2021

Вспоминаем линейную регрессию





Вспоминаем линейную регрессию

price	lotsize	bedrooms	bathrms	stories	driveway	recroom	fullbase	gashw	airco	garagepl	prefarea
42000.0	5850	3	1	2	yes	no	yes	no	no	1	no
38500.0	4000	2	1	1	yes	no	no	no	no	0	no
49500.0	3060	3	1	1	yes	no	no	no	no	0	no
60500.0	6650	3	1	2	yes	yes	no	no	no	0	no
61000.0	6360	2	1	1	yes	no	no	no	no	0	no

$$m = 11 \quad \hat{\mathbf{y}}(\mathbf{x}) = \sum_{i=0}^{m} w_i \mathbf{x}_i$$

$$\vec{\mathbf{y}} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nm} \end{pmatrix} \quad \vec{\mathbf{w}} = \begin{pmatrix} w_0 \\ \vdots \\ w_m \end{pmatrix}$$

$$\hat{\vec{\mathbf{y}}} = \mathbf{X}\vec{\mathbf{w}}$$

Вспоминаем линейную регрессию

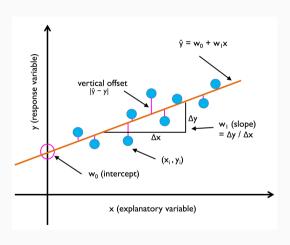
Обучение:

$$\mathcal{L}_{MSE}(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$$

$$\vec{\mathit{W}}^* = \underset{_{\mathit{W}}}{\mathsf{arg\,min}} \mathcal{L}$$

Аналитическое решение:

$$\vec{w}^* = (X^T X)^{-1} X^T y$$



$$\vec{W}^* = (X^T X)^{-1} X^T y$$

• Может ли быть так, что формула выше неприменима?

$$\vec{W}^* = (X^T X)^{-1} X^T y$$

- Может ли быть так, что формула выше неприменима?
- Может ли быть так, что никакой формулы нет?

$$\vec{W}^* = (X^T X)^{-1} X^T y$$

- Может ли быть так, что формула выше неприменима?
- Может ли быть так, что никакой формулы нет?
- Что в таком случае делать?

$$\vec{W}^* = (X^T X)^{-1} X^T y$$

- Может ли быть так, что формула выше неприменима?
- Может ли быть так, что никакой формулы нет?
- Что в таком случае делать?
- Наивный вариант: перебор по сетке
- $n = 100, m = 12 \rightarrow N_{ops} = 100^{12} = 10^{24}$

$$\vec{W}^* = (X^T X)^{-1} X^T y$$

- Может ли быть так, что формула выше неприменима?
- Может ли быть так, что никакой формулы нет?
- Что в таком случае делать?
- Наивный вариант: перебор по сетке
- $n = 100, m = 12 \rightarrow N_{ops} = 100^{12} = 10^{24}$
- Численные итеративные алгоритмы

Градиентный спуск



Метод обучения: градиентный спуск

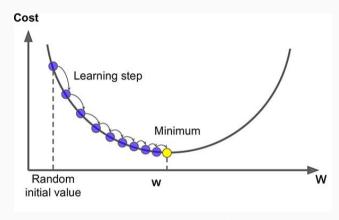
Итеративный алгоритм:

$$W_{i,t+1} = W_{i,t} - \alpha \frac{\partial \mathcal{L}_t}{\partial W_i}$$

В векторном виде:

$$\vec{\nabla} \mathcal{L}_t = \left(\frac{\partial \mathcal{L}_t}{\partial w_1}, \frac{\partial \mathcal{L}_t}{\partial w_2}, \cdots \right)$$

$$\vec{W}_{t+1} = \vec{W}_t - \alpha \vec{\nabla} \mathcal{L}_t$$



 α — learning rate, шаг обучения (гиперпараметр)



Биологические vs. искусственные нейроны







Искусственная птица

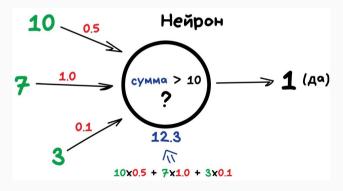
Биологические vs. искусственные нейроны



Искусственный нейрон

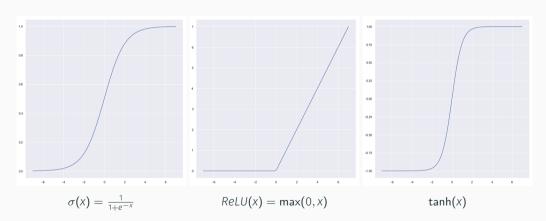
Матмодель нейрона: линейная регрессия + функция активации

$$n_{\text{out}} = f(w_0 + \sum_{i=1}^m w_i x_i)$$

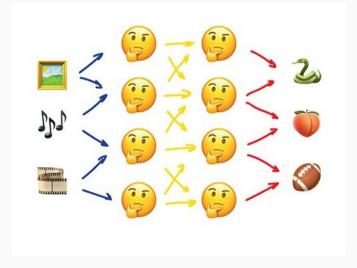


Функции активации

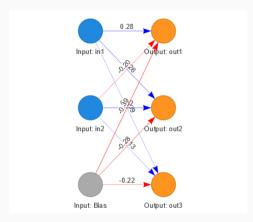
Функция активации — нелинейная функция, применяемая к выходу сумматора



Полносвязные нейронные сети



Нейронная сеть



Однослойный перцептрон

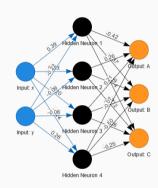
$$\vec{out} = \begin{pmatrix} out_1 \\ out_2 \\ out_3 \end{pmatrix} \quad X = \begin{pmatrix} 1 & in1_1 & in2_1 \\ \vdots & \vdots & \vdots \\ 1 & in1_N & in2_N \end{pmatrix}$$

$$W = \begin{pmatrix} b_1 & b_2 & b_3 \\ w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix}$$

$$\vec{out} = f(XW)$$

f применяется поэлементно

Полносвязная нейронная сеть



Сеть со скрытым слоем

$$\begin{aligned}
out &= \begin{pmatrix} A \\ B \\ C \end{pmatrix} & X &= \begin{pmatrix} 1 & x_1 & y_1 \\ \vdots & \vdots & \vdots \\ 1 & x_N & y_N \end{pmatrix} \\
W_1 &= \begin{pmatrix} b_1 & b_2 & b_3 & b_4 \\ w_{11} & w_{12} & w_{13} & w_{14} \\ w_{21} & w_{22} & w_{23} & w_{24} \end{pmatrix} & H &= f(XW_1) \\
W_2 &= \begin{pmatrix} b_1 & b_2 & b_3 \\ w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \end{pmatrix} \\
out &= f(HW_2) &= f(f(XW_1)W_2)
\end{aligned}$$

Полносвязная нейронная сеть

