« »

« »

:

:

204

..

19.05.2022

:

, ,

, ..-..

,

( )

Трушин

22 2022 9

(10-)

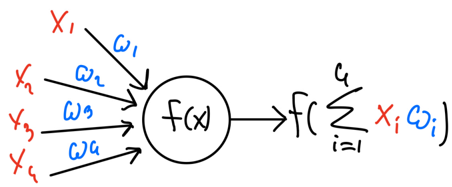
2022

.

1

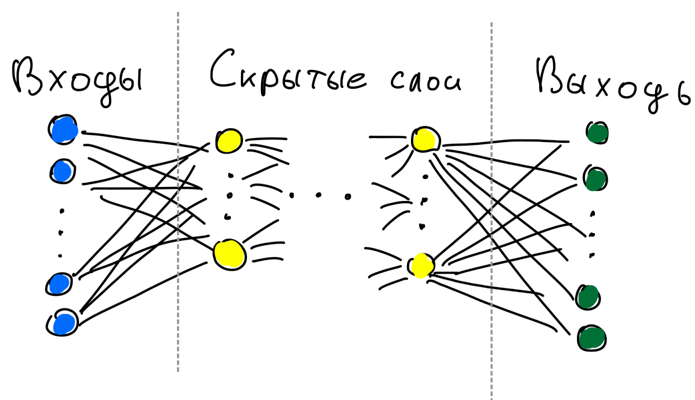
github.

– . , .
. , . . , , , , .
. , - .
– .



. 1:

, , , .



. 2:

. – . , .
, . . , , , . - , , .
. . m y_i, n , x_i. $y_i = f\left(\sum_{j=1}^{n} w_{ij}x_j + b_i\right)$, w_{ij} j- i- , b_i – , , f – . ,
$y = f(Ax+b)$, A – , b – , f . , $\theta = (A,b)$ f.

$$\mathbb{R}^n \xrightarrow{x} \underset{f(x,\theta)}{\boxed{\theta}} \xrightarrow{y} \mathbb{R}^m$$

, . k , $x = [x^{(1)}, \dots, x^{(k)}]^t$, y l $y = [y^{(1)}, \dots, y^{(l)}]^t$, . y x $F(x) = y$. , n , ,
. , F.
. θ_1, θ_2

$$\mathbb{R}^n \xrightarrow{x_i} \underset{f(x,\theta_1)}{\boxed{\theta_1}} \xrightarrow{w_i} \underset{g(x,\theta_2)}{\boxed{\theta_2}} \xrightarrow{z_i} \mathbb{R}^m$$

$\theta_1, \theta_2,$ $x_i \in \mathbb{R}^n, y_i \in \mathbb{R}^m$.

$$\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} \to \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

$\phi(\theta_1, \theta_2) = \sum_{i=1}^{p} \|g(f(x_i, \theta_1), \theta_2) - y_i\|^2 \to \underline{\min},$. θ_1, θ_2 .
, : , , . . . :

1.
2.
3.
4.
5. .

2

. : , , . , n . $y_i,$ $x_i = [x_i^{(1)}, x_i^{(2)}, x_i^{(3)}]^t$. f, $f(x_i) = y_i$. , , :

$$\sum_{i=0}^{n} |f(x_i) - y_i| \to \min$$

3

.

1. Net. , . , .
2. ComputeBlock. "". . , .
3. LosFunction. ,
4. ActivationFunction. , . , .
   (a) Sigmoid
   (b) Relu
   (c) Softmax

4

- C++20 [1]
- Google C++ Style Guide [2]
- ClangFormat linter [3]
- Eigen [4]
- : git [5] github [6]

5

5.1

$$\mathbb{R}^n \to \underset{f_1(x,\theta_1)}{\boxed{\theta_1}} \to \cdots \to \underset{f_i(x,\theta_i)}{\boxed{\theta_i}} \to \cdots \to \underset{f_k(x,\theta_k)}{\boxed{\theta_k}} \to \mathscr{L} \to \mathbb{R}$$

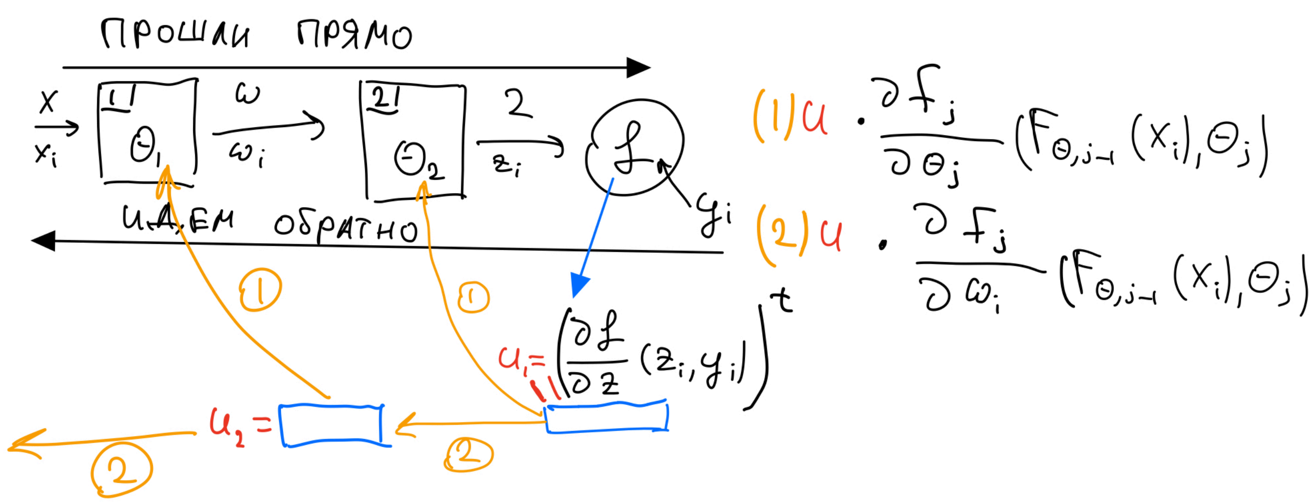
$f_i(x) = \phi(A_i x + b_i)$, $\phi(x)$ – , $(A_i, b_i) = \theta_i$ – , $\mathscr{L}$ – . $F_\Theta(x)$, $\Theta = (\theta_1, \dots, \theta_k)$, F_Θ i $i+1$, . , $F_{\Theta,i}(x) = f_i(F_{\Theta,i-1}(x), \theta_i)$, $F_{\Theta,1}(x) = f_1(x, \theta_1)$, $F_\Theta(x) = F_{\Theta,n}(x)$. $\psi(\Theta) = \frac{1}{n}\sum\limits_{i=1}^{n} \mathscr{L}(F_\Theta(x_i), y_i)$.
, , , , . $\frac{\partial \psi}{\partial \theta_j}$.

$$\frac{\partial \psi}{\partial \theta_j} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \mathscr{L}(F_\Theta(x_i), y_i)}{\partial F_\Theta(x_i)}\frac{\partial F_\Theta(x_i)}{\partial \theta_j}\right)$$

$\frac{\partial \mathscr{L}(z,y)}{\partial z}$ – .

$$\frac{\partial F_\Theta(x)}{\partial \theta_i} = \prod_{j=0}^{n-i+1}\left(\frac{\partial f_{n-j}(F_{\Theta,n-j-1}(x)), \theta_{n-j})}{\partial F_{\Theta,n-j-1}(x)}\right) \cdot \frac{\partial f_i(F_{\Theta,i-1}(x), \theta_i)}{\partial \theta_i}$$

, , , $\frac{\partial F_{\Theta,n}}{\partial \theta_n} = \frac{\partial f_n(F_{\Theta,n-1}(x), \theta_n)}{\partial \theta_n}$, $\prod\limits_{j=0}^{n-i+1}\left(\frac{\partial f_{n-j}(F_{\Theta,n-j-1}(x)), \theta_{n-j})}{\partial F_{\Theta,n-j-1}(x)}\right)$, . , $\frac{\partial F_{\Theta,n}}{\partial \theta_{n-1}} = \frac{f_n(F_{\Theta,n-1}(x), \theta_n)}{\partial F_{\Theta,n-1}(x)} \cdot \frac{\partial f_{n-1}(F_{\Theta,n-2}(x), \theta_{n-1})}{\partial \theta_{n-1}}$, , . , .



. 3:

. $u_1 = \left(\frac{\partial \mathscr{L}(z_i, y_i)}{\partial z}\right)^t$. θ_2 $u_1 \cdot \frac{\partial f_2(w_i, \theta_2)}{\partial \theta_2}$ $u_2 = u_1 \cdot \frac{\partial f_2(w_i, \theta_2)}{\partial w_i}$, u_2 .

$$u^t \frac{\partial f(x,\theta)}{\partial x}, u^t \frac{\partial f(x,\theta)}{\partial \theta} = \left(u^t \frac{\partial f(x,\theta)}{\partial A}, u^t \frac{\partial f(x,\theta)}{\partial b}\right)$$

$$f(x, A, b) = \phi(A_i x + b_i)$$

$$u^t d(\phi(Ax+b)) = u^t \phi'(Ax+b) d(Ax+b) = u^t \phi'(Ax+b) db = \langle \left(u^t \phi'(Ax+b)\right)^t, db \rangle \Rightarrow u^t \frac{\partial f(x,\theta)}{\partial b} = \phi'(Ax+b)^t u$$

$$u^t d(\phi(Ax+b)) = u^t \phi'(Ax+b) d(Ax+b) = u^t \phi'(Ax+b)(dA)x = \operatorname{tr}\left(u^t \phi'(Ax+b)(dA)x\right) = \operatorname{tr}\left(x u^t \phi'(Ax+b)(dA)\right) =$$

$$\circleddash\langle \left(xu^t\phi'(Ax+b)\right)^t, dA\rangle_F \Rightarrow u^t \frac{\partial f(x,\theta)}{\partial A} = \phi'(Ax+b)^t u x^t$$

$$u^t d(\phi(Ax+b)) = u^t\phi'(Ax+b)d(Ax+b) = u^t\phi' A dx = \langle (u^t\phi'(Ax+b)A)^t, dx\rangle \Rightarrow u^t \frac{\partial f(x,\theta)}{\partial x} = A^t\phi'(Ax+b)u$$

MSE, $\mathscr{L}(z,y) = \|z-y\|_2^2 = (z-y)^t(z-y)$

$$d\mathscr{L}(z,y) = 2(z-y)^t dx, \frac{\partial \mathscr{L}(z,y)}{\partial z} = 2(z-y)$$

.

sigmoid

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

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

, $\frac{\partial\sigma(x)}{\partial x}, x \in \mathbb{R}^n$, $\sigma'(x_i)$,
relu
" x

$$\text{relu}(x) = \begin{cases} x, & x > 0 \\ 0.01x, & x \leq 0 \end{cases}$$

$$\text{relu}'(x) = \begin{cases} 1, & x > 0 \\ 0.01, & x \leq 0 \end{cases}$$

sigmoid'e, $\frac{\partial\,\text{relu}(x)}{\partial x}, x \in \mathbb{R}^n$, $\text{relu}'(x_i)$,
softmax

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum\limits_{j=1}^{n} e^{x_j}}, x \in \mathbb{R}^n$$

$$\frac{\partial\,\text{softmax}(x)}{\partial x} = \begin{bmatrix} \frac{\partial s_1}{\partial x_1} & \frac{\partial s_1}{\partial x_2} & \cdots & \frac{\partial s_1}{\partial x_n} \\ \frac{\partial s_2}{\partial x_1} & \frac{\partial s_2}{\partial x_2} & \cdots & \frac{\partial s_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial s_n}{\partial x_1} & \frac{\partial s_n}{\partial x_2} & \cdots & \frac{\partial s_n}{\partial x_n} \end{bmatrix}, s_i = \text{softmax}(x)_i$$

, , $\frac{\partial \mathscr{L}(F_\Theta(x_i), y_i)}{\partial\theta_j}$, x_i, $\frac{\partial\psi}{\partial\theta_i}$ (learning rate), ,

$$\theta_i' = \theta_i - lr \cdot \frac{\partial\psi}{\partial\theta_i}$$

5.2 Net

. . . . .

- , , ,
- train – . x – , y – ,
- predict_1d –
- predict_2d – ,
- push_forward – ,
- back_propagate – , ,
- update_parameters – ,

5.3 ComputeBlock

. $\theta = (A, b)$ ϕ. $f(x) = \phi(Ax + b)$, $\frac{\partial f}{\partial A}, \frac{\partial f}{\partial b}, \frac{\partial f}{\partial x}$.

- , A b $[-1, 1]$
- evaluate_1d – $f(x) = \phi(Ax + b)$
- evaluate_2d – 1d, ,
- push_forward –
- back_propagate –
- update_parameters –
- grad_A – $\frac{\partial f}{\partial A}$
- grad_b – $\frac{\partial f}{\partial b}$
- grad_x – $\frac{\partial f}{\partial x}$

5.4

softmax, relu, sigmoid evaluate derivative,

5.5

MSE

- evaluate_1d – z, y
- evaluate_2d – z, y
- grad_z – $\frac{\partial \mathscr{L}(z,y)}{\partial z}$

6

, .

mnist [7], 90%.

- 8500
- 784 , 28×28, relu
- 16 , relu
- 16 , softmax
- 10
- ( ) = 3000
- Learning rate ( ) = 0.6
- 128
- MSE

128 . relu, sigmoid . softmax, "". 10 "" , . CPU 12 . 1500 , 1351 , 90%.

[1] URL: https://en.cppreference.com/w/cpp/20.

[2] URL: https://google.github.io/styleguide/cppguide.html.

[3] URL: https://clang.llvm.org/docs/ClangFormat.html.

[4] URL: https://eigen.tuxfamily.org/index.php?title=Main_Page.

[5] URL: https://git-scm.com.

[6] URL: https://github.com.

[7] URL: http://yann.lecun.com/exdb/mnist/.