

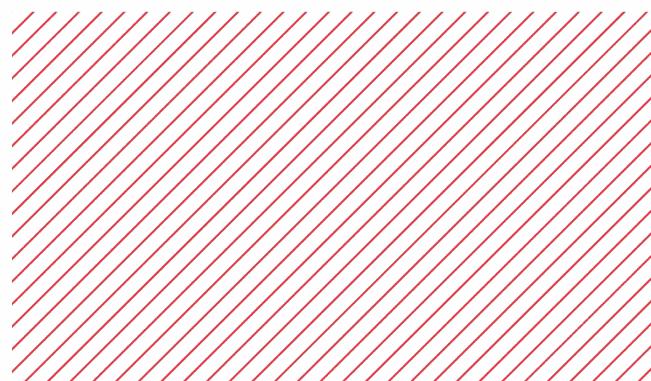
академия
больших
данных



Полносвязные сети. Метод обратного распространения ошибки

Фёдор Киташов

Программист-исследователь в
команде компьютерного зрения





План лекции

- Повторение: логистическая регрессия через матричные перемножения
- Полносвязанные сети
- Обучение полносвязных сетей через метод обратного распространения ошибки (backpropagation)

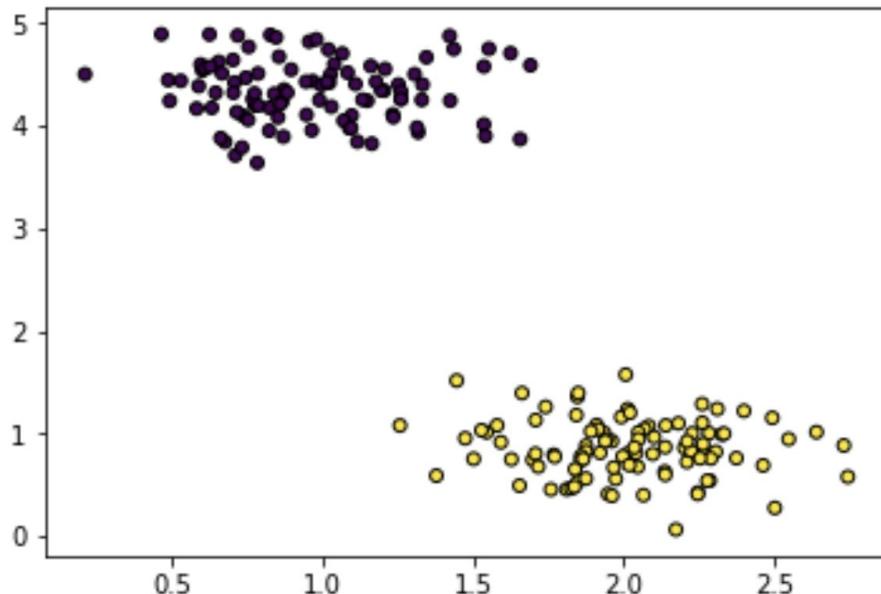
Повторение

```
import sklearn
from sklearn.datasets import make_blobs

X, y = make_blobs(n_samples=200, centers=2, n_features=2,
                  random_state=0, cluster_std=0.3)

plt.scatter(X[:, 0], X[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```



Повторение

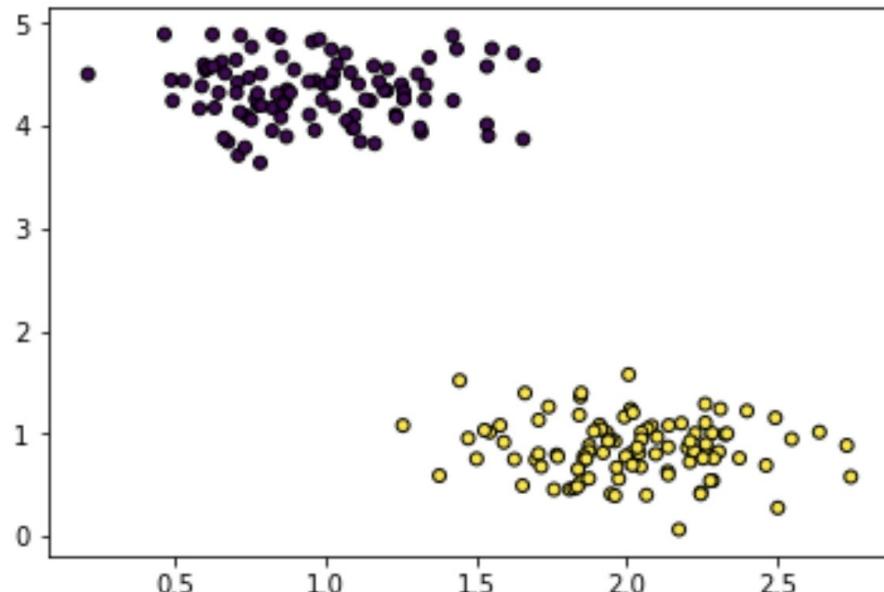
- Какие параметры модели?
- Какая размерность параметров модели?

```
import sklearn
from sklearn.datasets import make_blobs

X, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(X[:, 0], X[:, 1], marker='o', c=y,
            s=25, edgecolor='k')
```

<matplotlib.collections.PathCollection at 0x1a2969dac8>



Повторение

- Какие параметры модели?
- Какая размерность параметров модели?

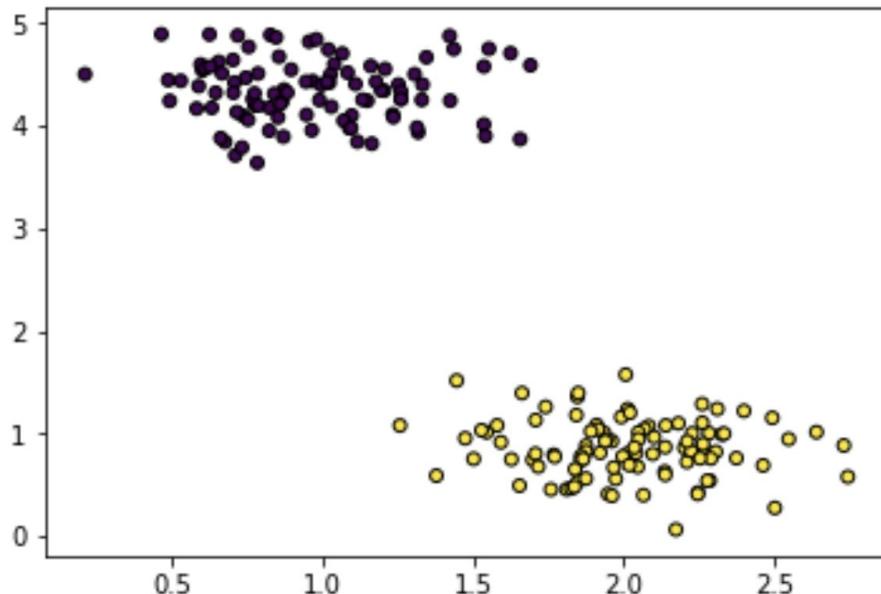
$$\begin{bmatrix} x \\ y \end{bmatrix}$$

```
import sklearn
from sklearn.datasets import make_blobs

X, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(X[:, 0], X[:, 1], marker='o', c=y,
            s=25, edgecolor='k')
```

<matplotlib.collections.PathCollection at 0x1a2969dac8>



Повторение

- Какие параметры модели?
- Какая размерность параметров модели?

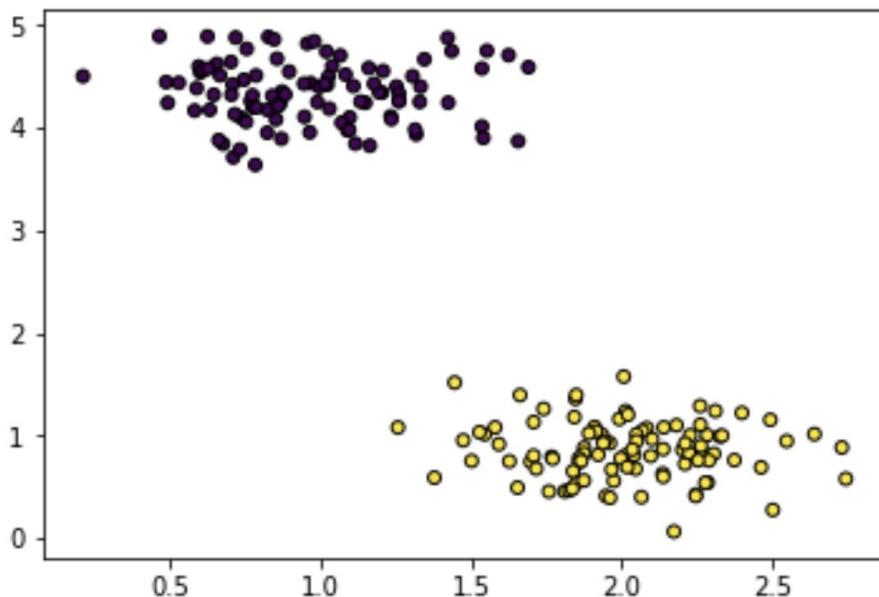
$$\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix}$$

```
import sklearn
from sklearn.datasets import make_blobs

X, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(X[:, 0], X[:, 1], marker='o', c=y,
            s=25, edgecolor='k')
```

<matplotlib.collections.PathCollection at 0x1a2969dac8>



Повторение

- Какие параметры модели?
- Какая размерность параметров модели?

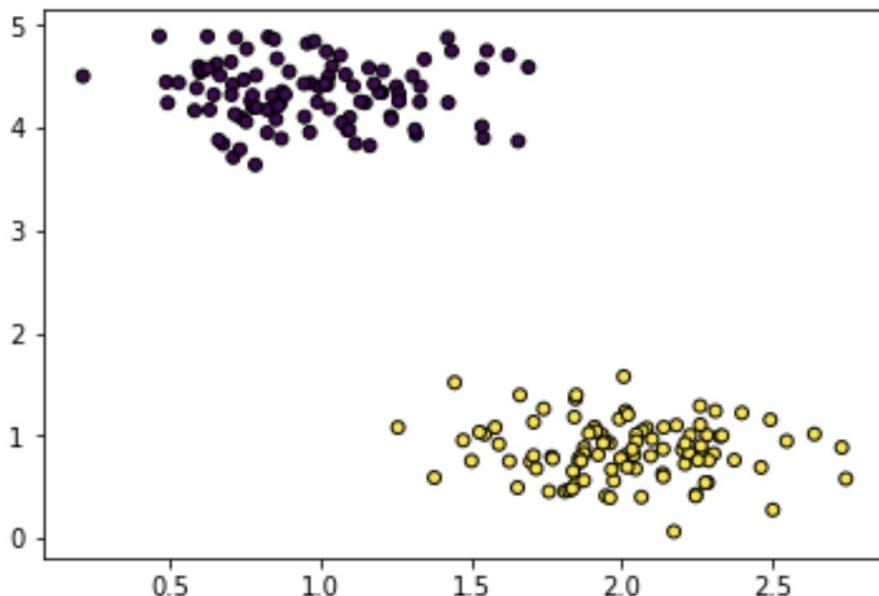
$$\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_1 \end{bmatrix}$$

```
import sklearn
from sklearn.datasets import make_blobs

x, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(x[:, 0], x[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```



Повторение

- Какие параметры модели?
- Какая размерность параметров модели?

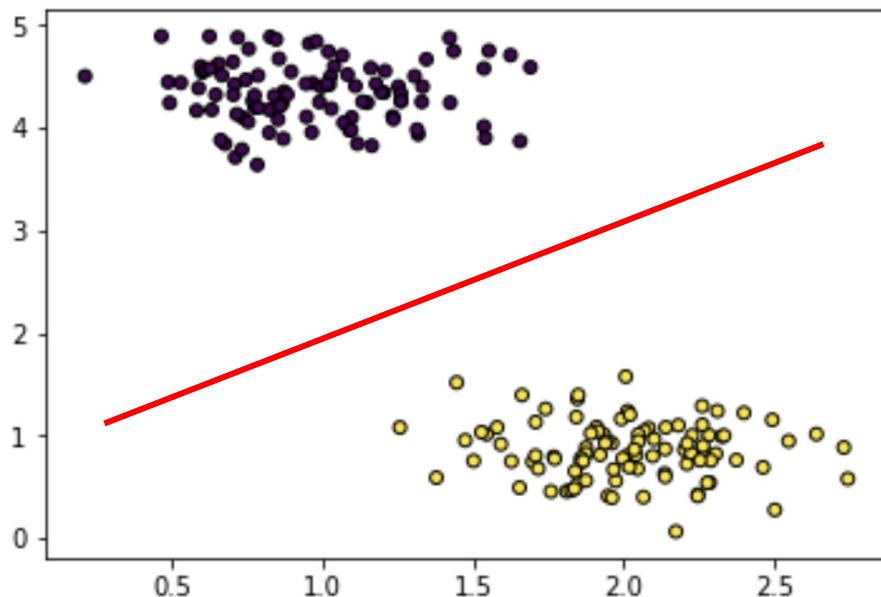
$$\text{sign}\left(\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}\right)$$

```
import sklearn
from sklearn.datasets import make_blobs

x, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(x[:, 0], x[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```



Повторение: логистическая регрессия

- Какие параметры модели?
- Какая размерность параметров модели?

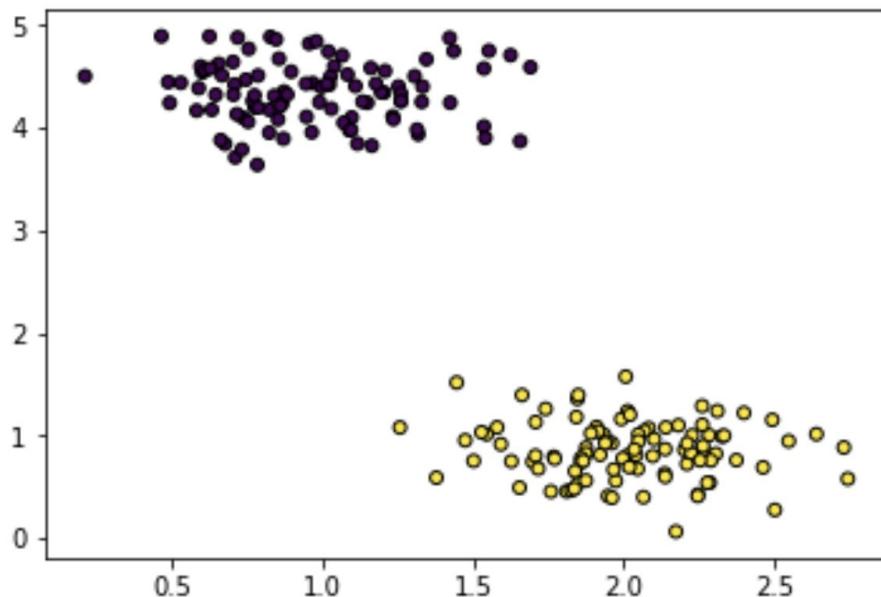
$$p = \text{sigmoid} \left(\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \right)$$

```
import sklearn
from sklearn.datasets import make_blobs

x, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(x[:, 0], x[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```



Повторение: логистическая регрессия

- Какие параметры модели?
- Какая размерность параметров модели?

$$p = \text{sigmoid} \left(\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \right)$$

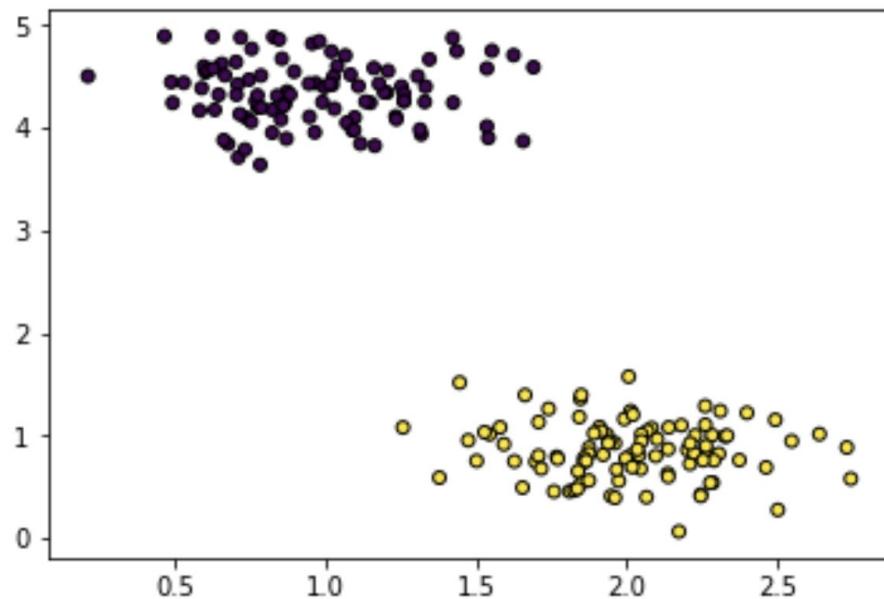
$$J = -(y \log(p) + (1 - y) \log(1 - p))$$

```
import sklearn
from sklearn.datasets import make_blobs

x, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(x[:, 0], x[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```



Повторение: логистическая регрессия

- Какие параметры модели?
- Какая размерность параметров модели?
- **Как обновлять веса?**

$$p = \text{sigmoid} \left(\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \right)$$

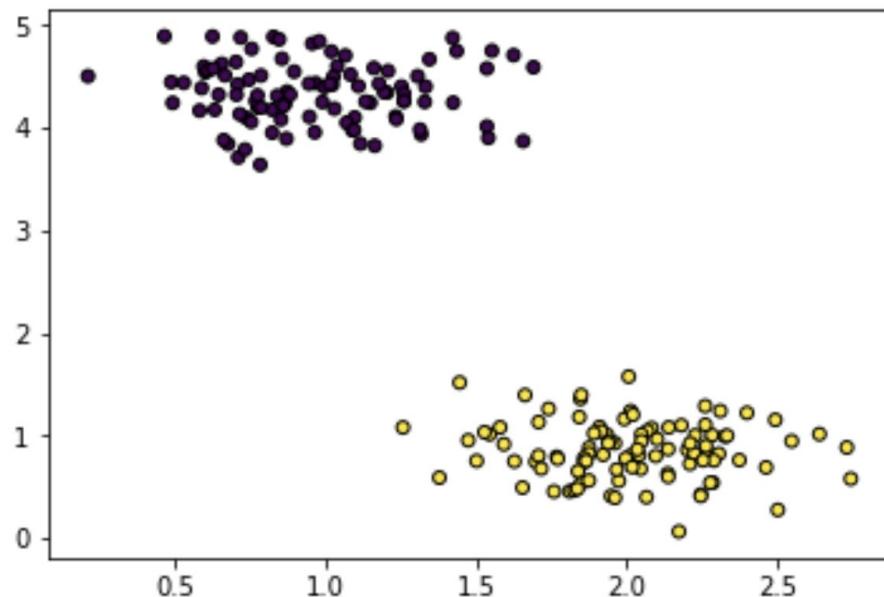
$$J = -(y \log(p) + (1 - y) \log(1 - p))$$

```
import sklearn
from sklearn.datasets import make_blobs

x, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(x[:, 0], x[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```



Повторение: логистическая регрессия

- Какие параметры модели?
- Какая размерность параметров модели?
- **Как обновлять веса?**

$$p = \text{sigmoid} \left(\begin{bmatrix} w_0 & w_1 \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \right)$$

$$J = -(y \log(p) + (1 - y) \log(1 - p))$$

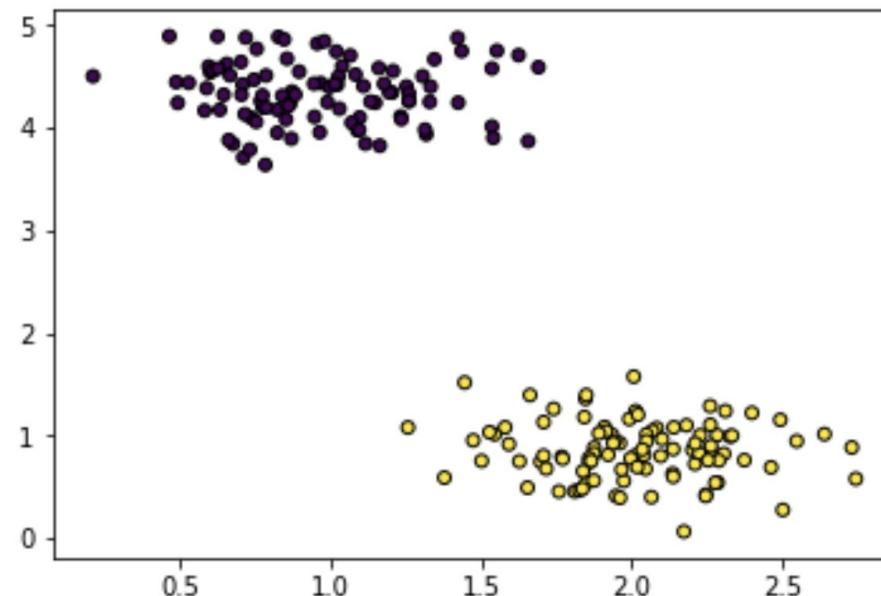
$$W^{t+1} = W^t - \eta \times \nabla Loss_W$$

```
import sklearn
from sklearn.datasets import make_blobs

x, y = make_blobs(n_samples=200, centers=2, n_features=2,
                   random_state=0, cluster_std=0.3)

plt.scatter(x[:, 0], x[:, 1], marker='o', c=y,
            s=25, edgecolor='k')

<matplotlib.collections.PathCollection at 0x1a2969dac8>
```





Проблемы

- Как быть, если классов больше, чем 2?
- Как быть, если проблема нелинейная?



Проблемы

- **Как быть, если классов больше, чем 2?**
- Как быть, если проблема нелинейная?



Многоклассовая классификация

$$p(y = 1 \mid x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$



Многоклассовая классификация

$$p(y = 1 \mid x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два?



Многоклассовая классификация

$$p(y = 1 \mid x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

Многоклассовая классификация

$$p(y = 1 | x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

$$\begin{bmatrix} x \\ y \end{bmatrix}$$

Многоклассовая классификация

$$p(y = 1 | x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

$$\begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix}$$

Многоклассовая классификация

$$p(y = 1 | x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

$$\begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix}$$

Многоклассовая классификация

$$p(y = 1 | x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

$$\begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix}$$

Многоклассовая классификация

$$p(y = 1 | x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

$$\text{softmax} \left(\begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \end{bmatrix} \times \begin{bmatrix} x \\ y \\ \vdots \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \right) = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix}$$

Многоклассовая классификация

$$p(y = 1 | x) = \frac{1}{1 + \exp(-\langle w, x \rangle)}$$

- Как быть, если классов больше, чем два? Предположим, что классов 3

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

σ = softmax

\vec{z} = input vector

e^{z_i} = standard exponential function for input vector

K = number of classes in the multi-class classifier

e^{z_j} = standard exponential function for output vector

$$\text{softmax} \left(\begin{bmatrix} w_{00} & w_{01} \\ w_{10} & w_{11} \\ w_{20} & w_{21} \end{bmatrix} \times \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} \right) = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix}$$



Softmax

```
import numpy as np

def softmax(xs):
    return np.exp(xs) / sum(np.exp(xs))

xs = np.array([-1, 0, 3, 5])
print(softmax(xs)) # [0.0021657, 0.00588697, 0.11824302, 0.87370431]
```



Проблемы

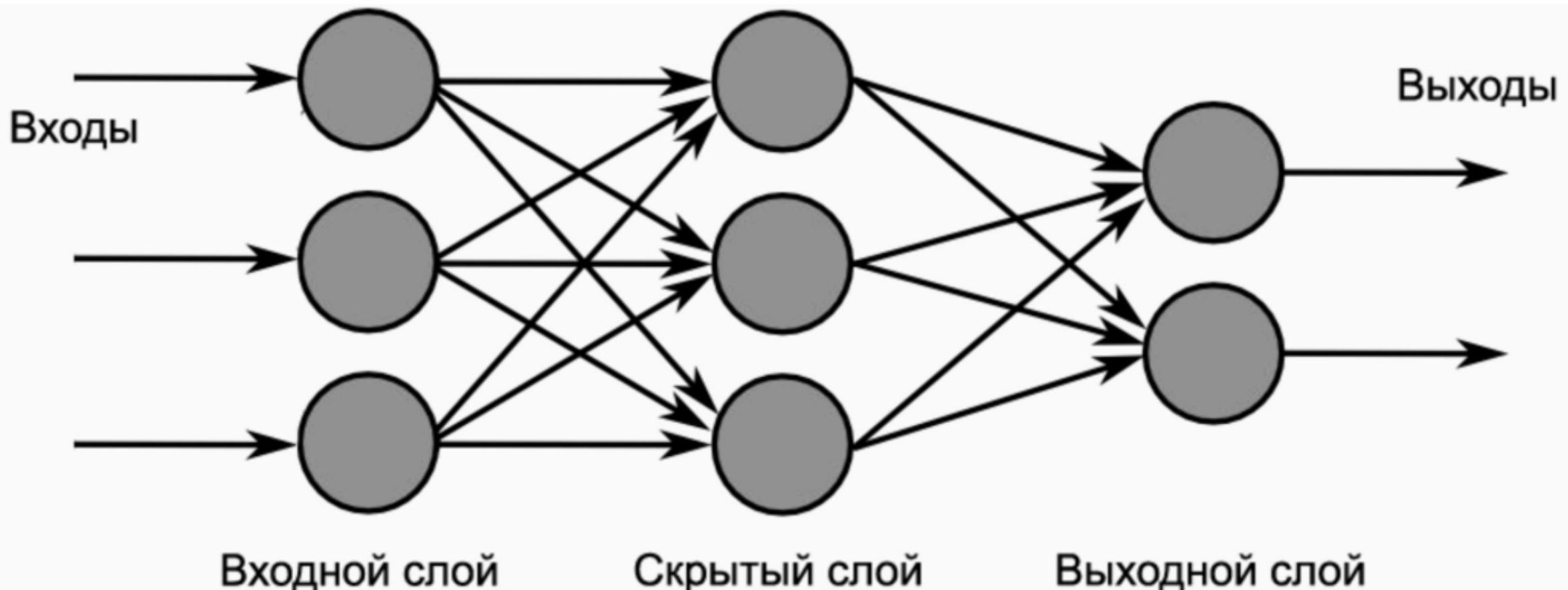
- Как быть, если классов больше, чем 2?
- **Как быть, если проблема нелинейная?**



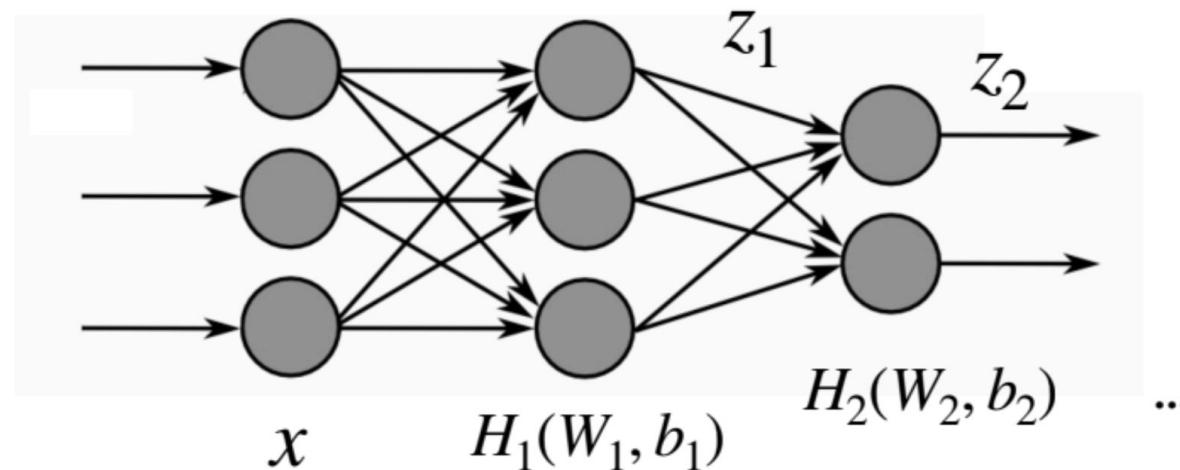
Как быть, если проблема нелинейная?

Интерактивная визуализация

Решение: многослойная сеть



Решение: многослойная сеть



$$z_1 = \text{Activation}(H_1(x)) = \text{Activation}(W_1 \times x + b_1)$$

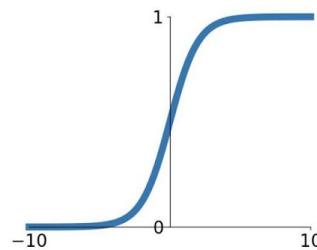
$$z_2 = \text{Activation}(H_2(z_1)) = \text{Activation}(W_2 \times z_1 + b_2)$$

$$W^{t+1} = W^t - \eta \times \nabla Loss_W$$

Активации a.k.a нелинейности

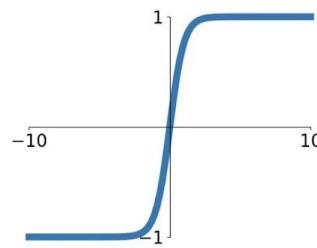
Sigmoid

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



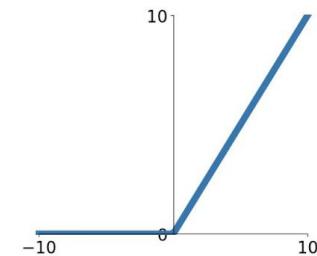
tanh

$$\tanh(x)$$



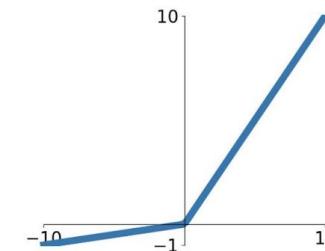
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

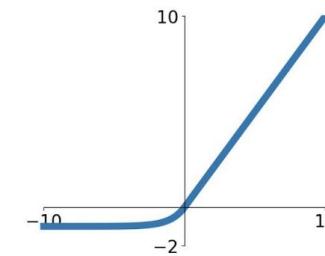


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$





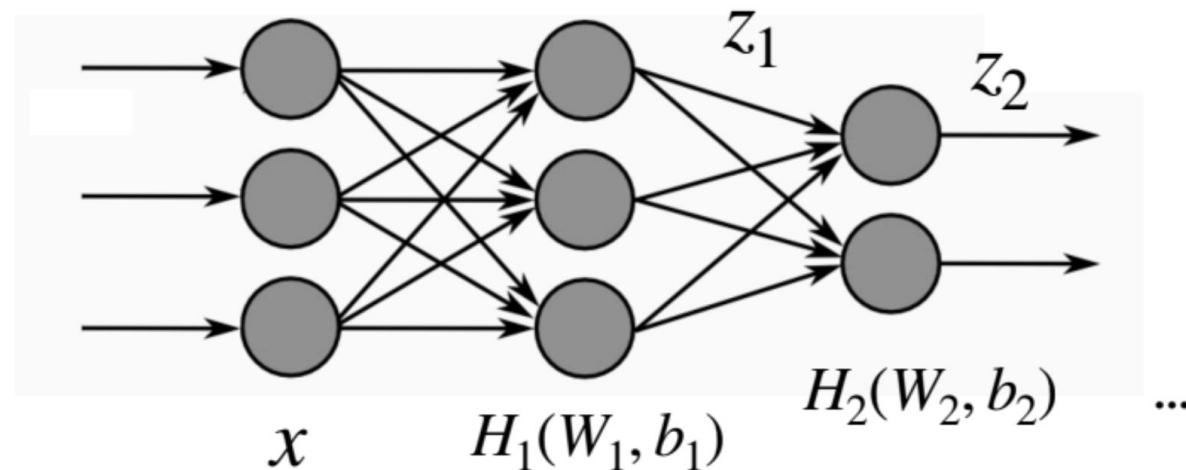
Какой лосс используется при классификации?



Какой лосс используется при классификации?

$$H(p, q) = - \sum_x p(x) \log q(x)$$

Обучение многослойной сети



$$z_1 = \text{Activation}(H_1(x)) = \text{Activation}(W_1 \times x + b_1)$$

$$z_2 = \text{Activation}(H_2(z_1)) = \text{Activation}(W_2 \times z_1 + b_2)$$

$$W^{t+1} = W^t - \eta \times \nabla Loss_W$$



Метод обратного распространения ошибки

$$x = \text{input}$$

$$z = Wx + b_1$$

$$h = \text{ReLU}(z)$$

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$



Метод обратного распространения ошибки

$$\mathbf{x} = \text{input}$$

$$\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}_1$$

$$\mathbf{h} = \text{ReLU}(\mathbf{z})$$

$$\boldsymbol{\theta} = \mathbf{U}\mathbf{h} + \mathbf{b}_2$$

$$\hat{\mathbf{y}} = \text{softmax}(\boldsymbol{\theta})$$

$$J = CE(\mathbf{y}, \hat{\mathbf{y}})$$

Какие градиенты нам нужно посчитать, чтобы обновить веса слоёв?



Метод обратного распространения ошибки

x = input

$z = Wx + b_1$

$h = \text{ReLU}(z)$

$\theta = Uh + b_2$

$\hat{y} = \text{softmax}(\theta)$

$J = CE(y, \hat{y})$

Какие градиенты нам нужно посчитать, чтобы обновить веса слоёв?

$$\frac{\partial J}{\partial U}$$

$$\frac{\partial J}{\partial b_2}$$

$$\frac{\partial J}{\partial W}$$

$$\frac{\partial J}{\partial b_1}$$

Метод обратного распространения ошибки

$$\mathbf{x} = \text{input}$$

$$\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}_1$$

$$\mathbf{h} = \text{ReLU}(\mathbf{z})$$

$$\boldsymbol{\theta} = \mathbf{U}\mathbf{h} + \mathbf{b}_2$$

$$\hat{\mathbf{y}} = \text{softmax}(\boldsymbol{\theta})$$

$$J = CE(\mathbf{y}, \hat{\mathbf{y}})$$

Какие градиенты нам нужно посчитать, чтобы обновить веса слоёв?

$$\frac{\partial J}{\partial \mathbf{U}}$$

$$\frac{\partial J}{\partial \mathbf{b}_2}$$

$$\frac{\partial J}{\partial \mathbf{W}}$$

$$\frac{\partial J}{\partial \mathbf{b}_1}$$

Какие эффективно посчитать эти градиенты?

Ответ: chain rule

Метод обратного распространения ошибки

x = input

$z = Wx + b_1$

$h = \text{ReLU}(z)$

$\theta = Uh + b_2$

$\hat{y} = \text{softmax}(\theta)$

$J = CE(y, \hat{y})$

Какие градиенты нам нужно посчитать, чтобы обновить веса слоёв?

$$\frac{\partial J}{\partial U}$$

$$\frac{\partial J}{\partial b_2}$$

$$\frac{\partial J}{\partial W}$$

$$\frac{\partial J}{\partial b_1}$$

Какие эффективно посчитать эти градиенты?

Ответ: chain rule

$$\frac{\partial J}{\partial U} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \frac{\partial \theta}{\partial U}$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \frac{\partial \theta}{\partial b_2}$$



Chain rule: $\frac{d}{dx} [f(g(x))] = f'(g(x))g'(x)$



Chain rule: $\frac{d}{dx} [f(g(x))] = f'(g(x))g'(x)$

$$h(x) = \underbrace{(\overbrace{5 - 6x}^{\text{inner}})^5}_{\text{outer}}$$

$$g(x) = 5 - 6x \quad \text{inner function}$$

$$f(x) = x^5 \quad \text{outer function}$$


$$\text{Chain rule: } \frac{d}{dx} [f(g(x))] = f'(g(x)) \cdot g'(x)$$

$$h(x) = (\underbrace{5 - 6x}_{\text{outer}})^5$$

$$g(x) = 5 - 6x \quad \text{inner function}$$

$$f(x) = x^5 \quad \text{outer function}$$


$$\text{Chain rule: } \frac{d}{dx} [f(g(x))] = f'(g(x)) \cdot g'(x)$$

$$h(x) = (\underbrace{5 - 6x}_{\text{outer}})^5$$

$$g(x) = 5 - 6x \quad \text{inner function}$$

$$f(x) = x^5 \quad \text{outer function}$$

$$g'(x) = -6$$

$$f'(x) = 5x^4$$


$$\text{Chain rule: } \frac{d}{dx} [f(g(x))] = f'(g(x)) \cdot g'(x)$$

$$h(x) = (\underbrace{5 - 6x}_{\text{outer}})^5$$

$$g(x) = 5 - 6x \quad \text{inner function}$$

$$f(x) = x^5 \quad \text{outer function}$$

$$g'(x) = -6$$

$$f'(x) = 5x^4$$

$$\frac{d}{dx} [f(g(x))]$$

$$= f'(g(x)) \cdot g'(x)$$

$$= 5(5 - 6x)^4 \cdot -6$$

$$= -30(5 - 6x)^4$$

Метод обратного распространения ошибки

x = input

$z = Wx + b_1$

$h = \text{ReLU}(z)$

$\theta = Uh + b_2$

$\hat{y} = \text{softmax}(\theta)$

$J = CE(y, \hat{y})$

Какие градиенты нам нужно посчитать, чтобы обновить веса слоёв?

$$\frac{\partial J}{\partial U}$$

$$\frac{\partial J}{\partial b_2}$$

$$\frac{\partial J}{\partial W}$$

$$\frac{\partial J}{\partial b_1}$$

Какие эффективно посчитать эти градиенты?

Ответ: chain rule

$$\frac{\partial J}{\partial U} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \frac{\partial \theta}{\partial U}$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \frac{\partial \theta}{\partial b_2}$$



Метод обратного распространения ошибки: одномерный случай

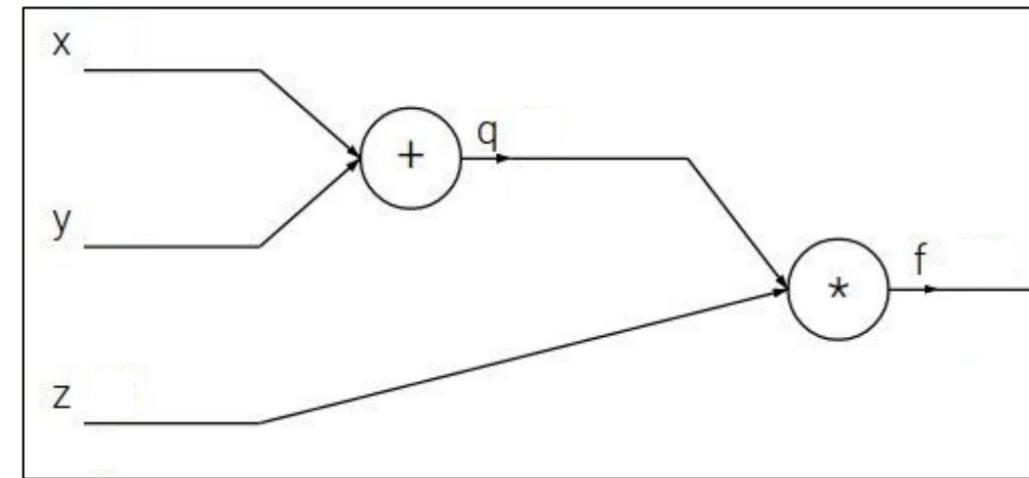
Backpropagation: a simple example

$$f(x, y, z) = (x + y)z$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

$$f(x, y, z) = (x + y)z$$

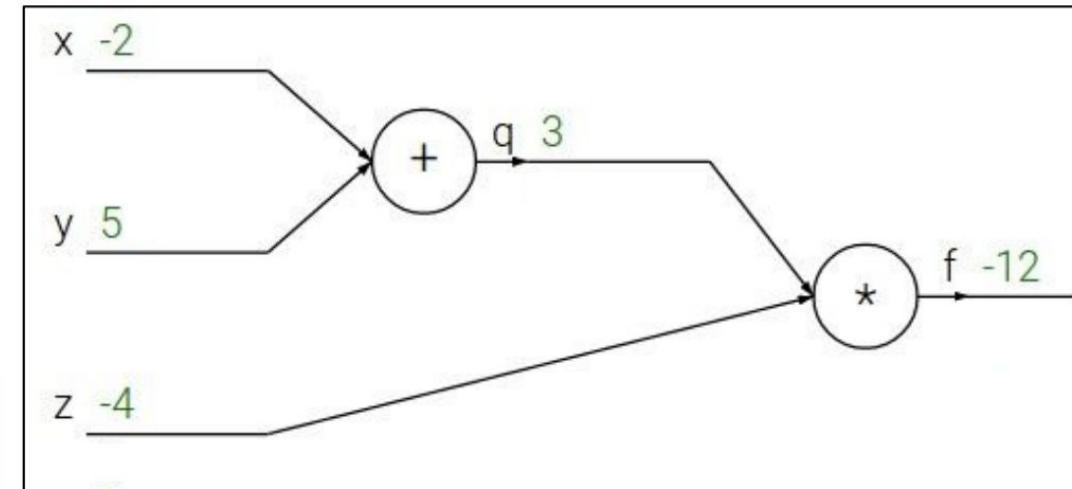


Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$



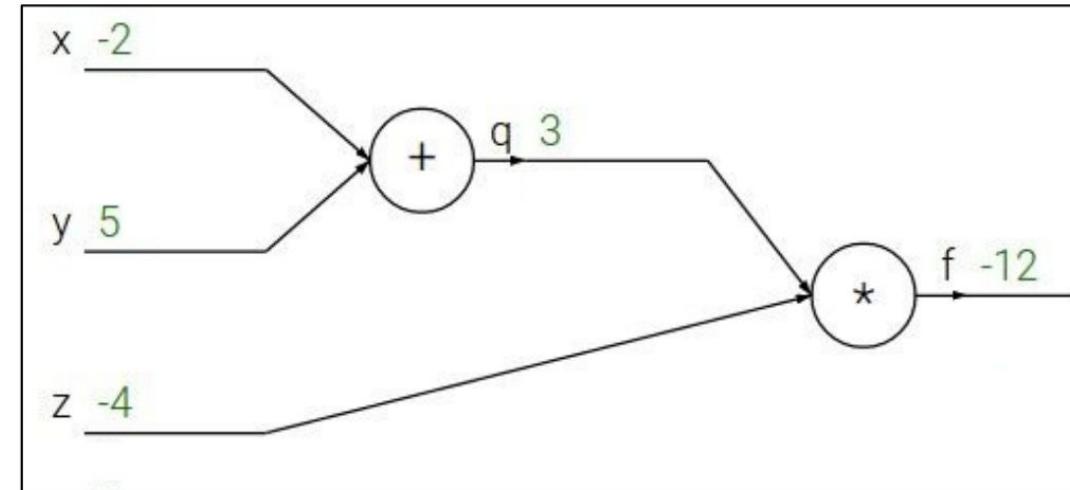
Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$



Метод обратного распространения ошибки: одномерный случай

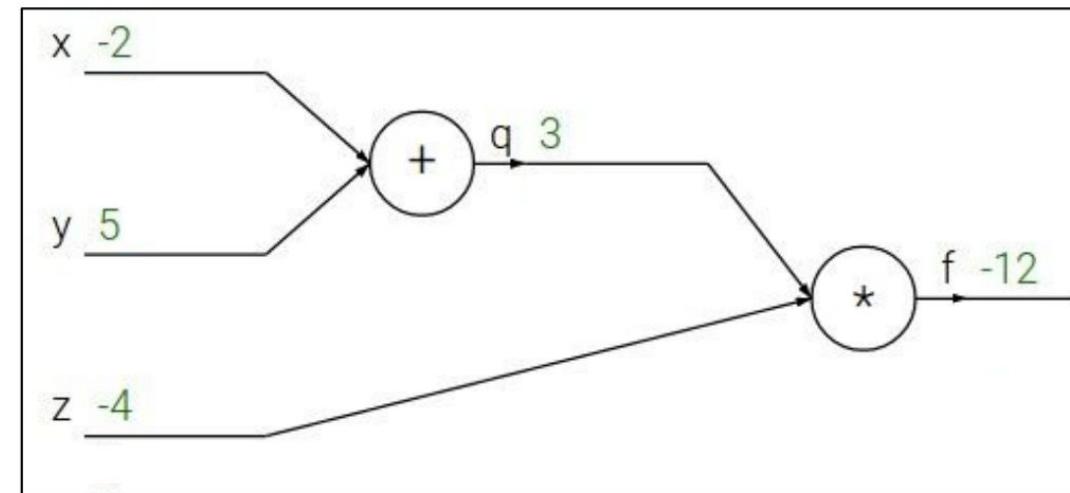
Backpropagation: a simple example

$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$



Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

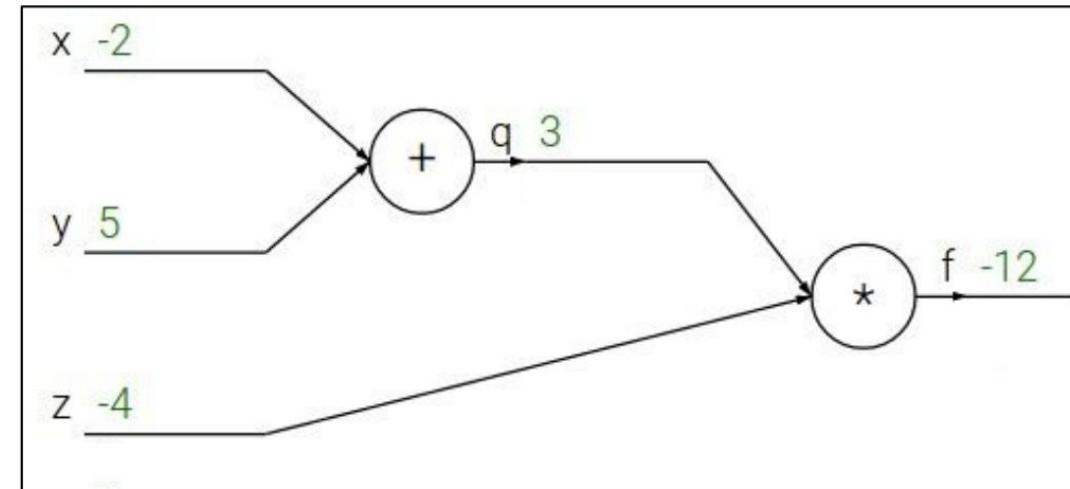
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

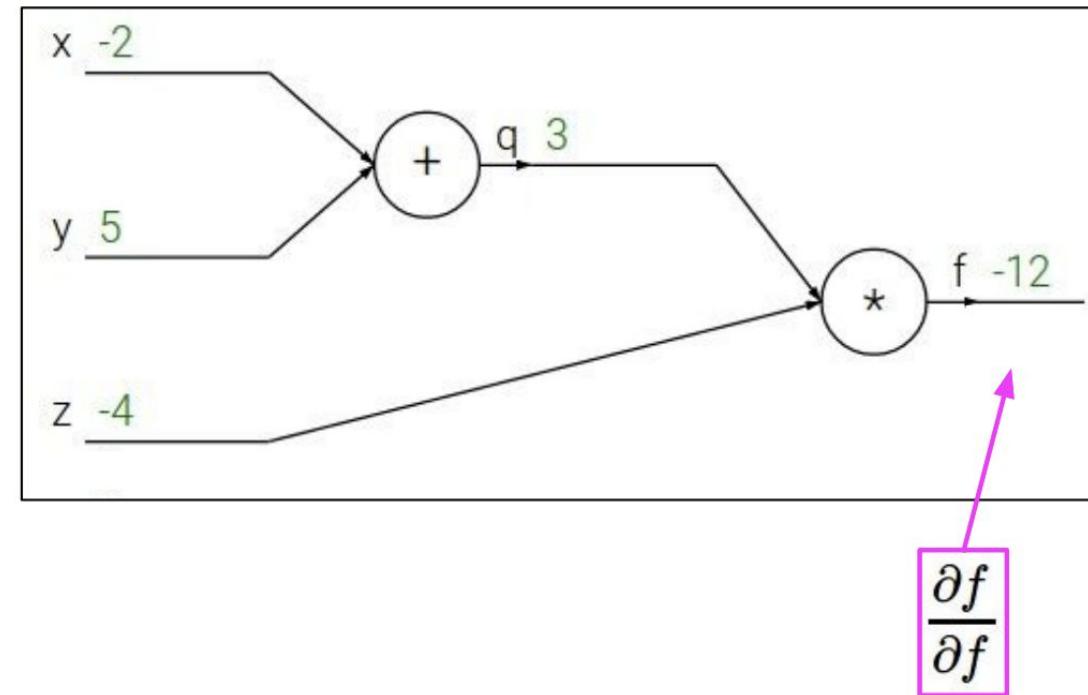
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

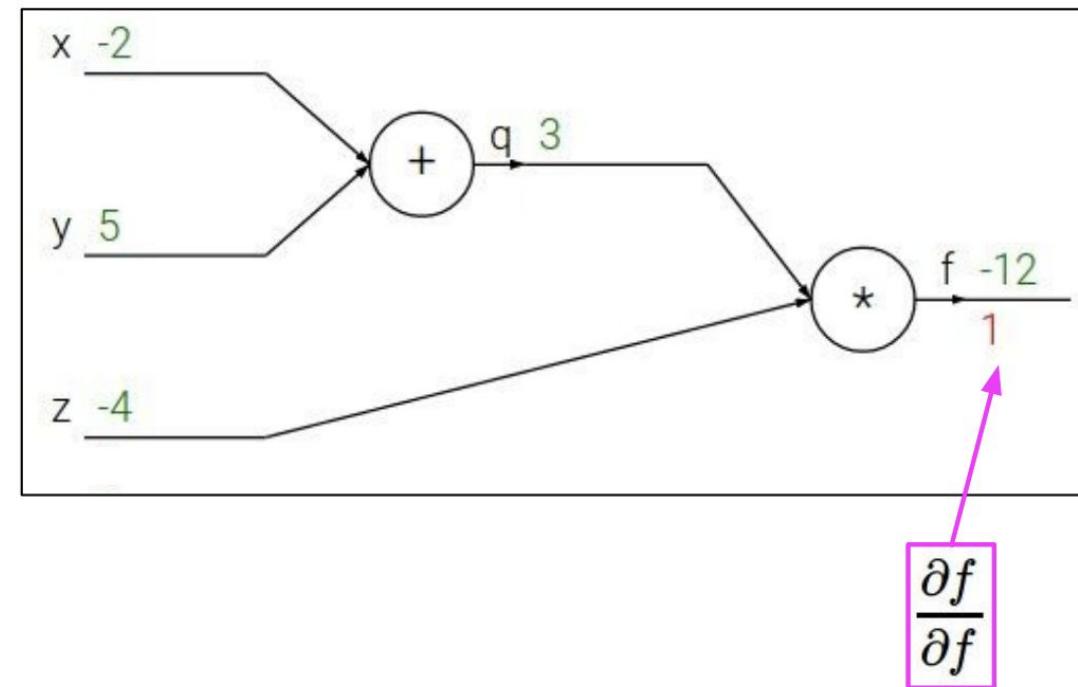
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2, y = 5, z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

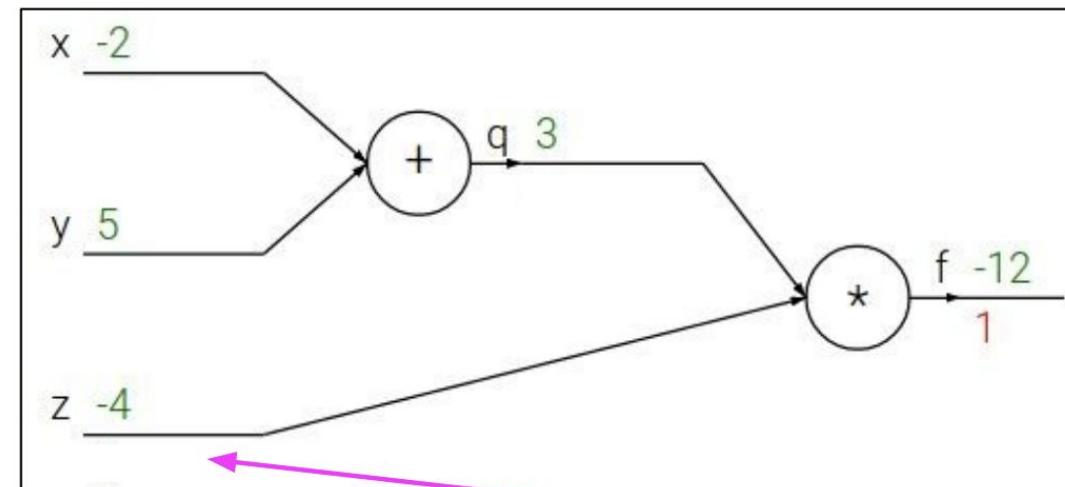
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial z}$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

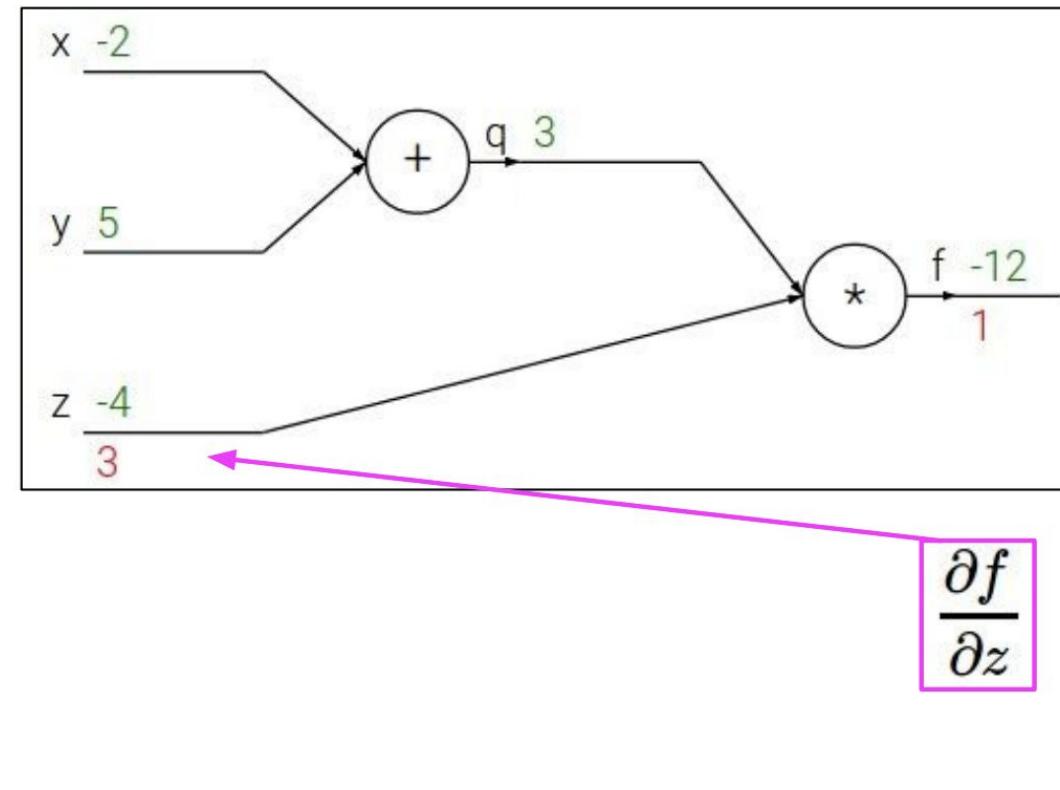
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

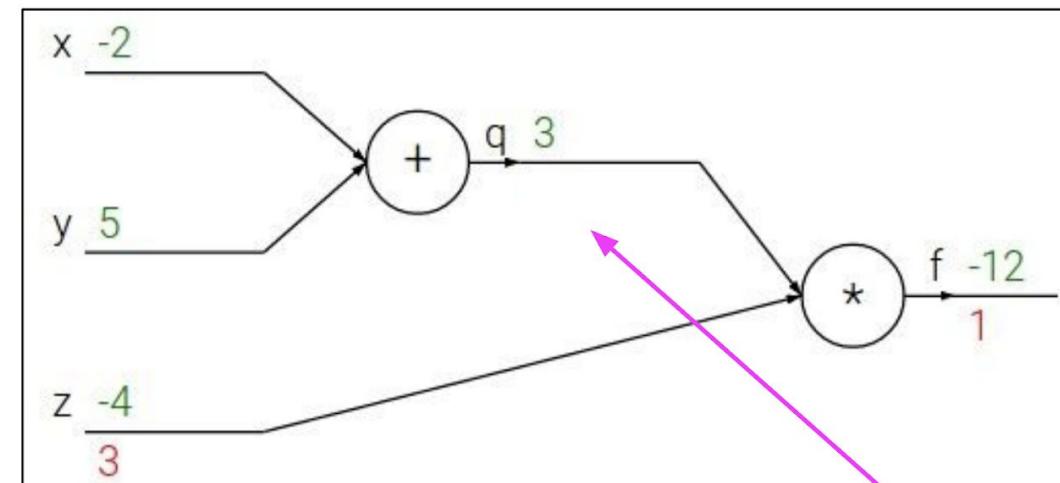
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial q}$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

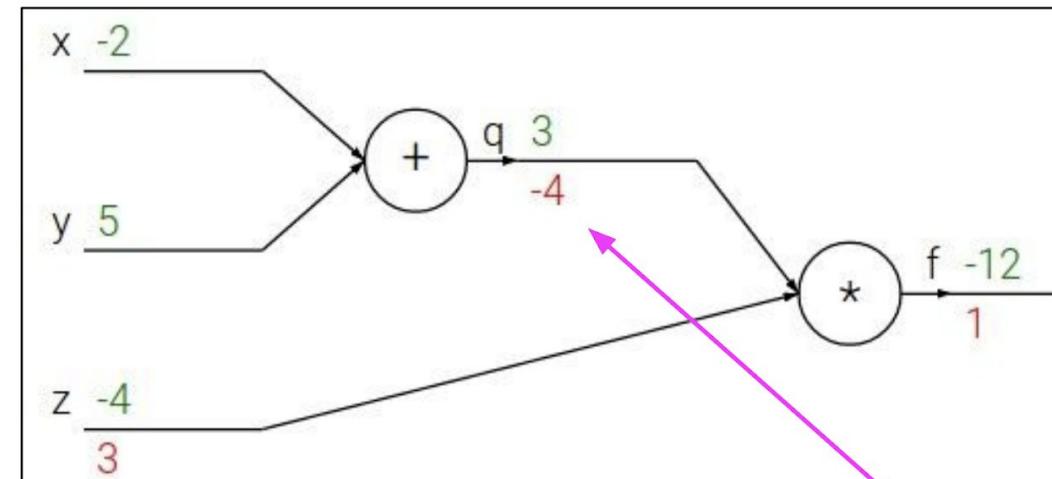
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



$$\frac{\partial f}{\partial q}$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

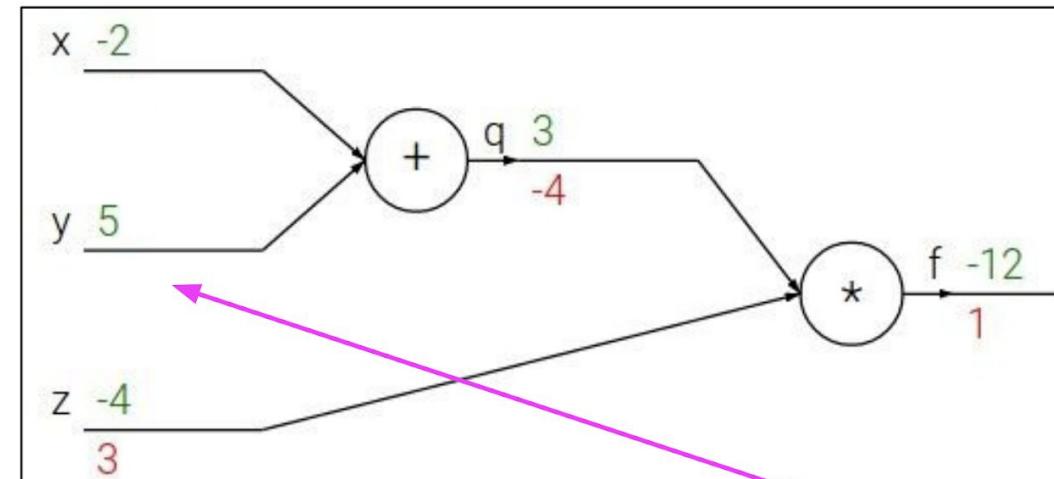
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Chain rule:

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial y}$$

Upstream gradient Local gradient

$$\frac{\partial f}{\partial y}$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

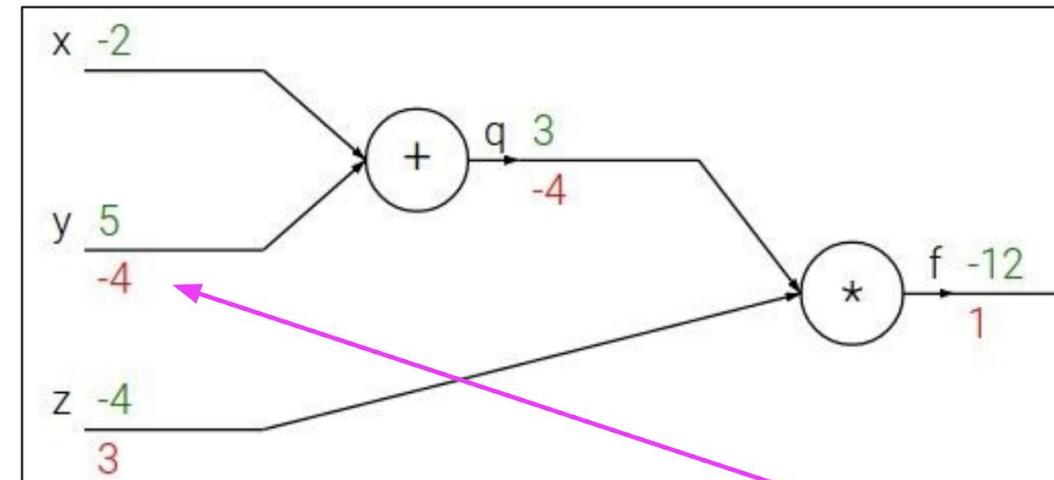
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Chain rule:

$$\frac{\partial f}{\partial y} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial y}$$

Upstream gradient Local gradient

$$\frac{\partial f}{\partial y}$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

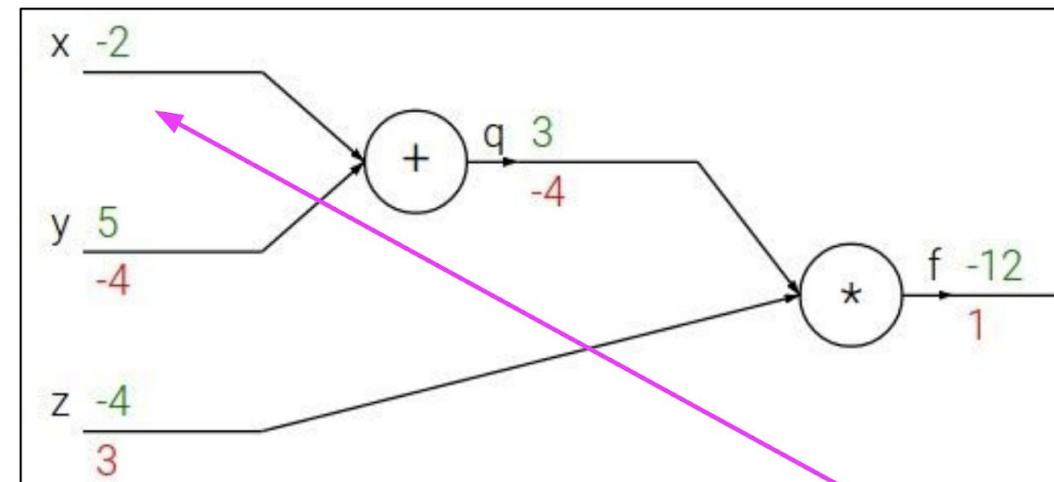
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Chain rule:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

Upstream gradient Local gradient

$$\frac{\partial f}{\partial x}$$

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

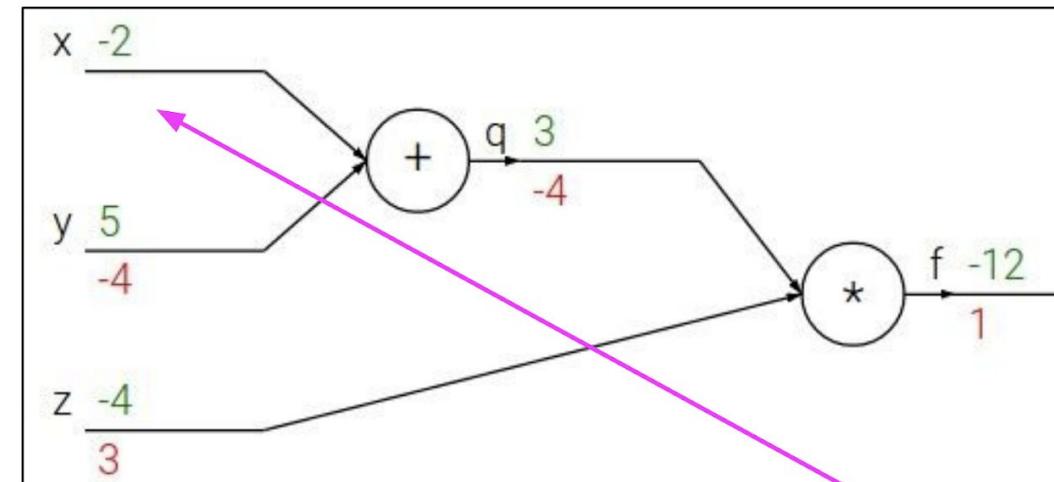
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Chain rule:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

Upstream gradient Local gradient

Метод обратного распространения ошибки: одномерный случай

Backpropagation: a simple example

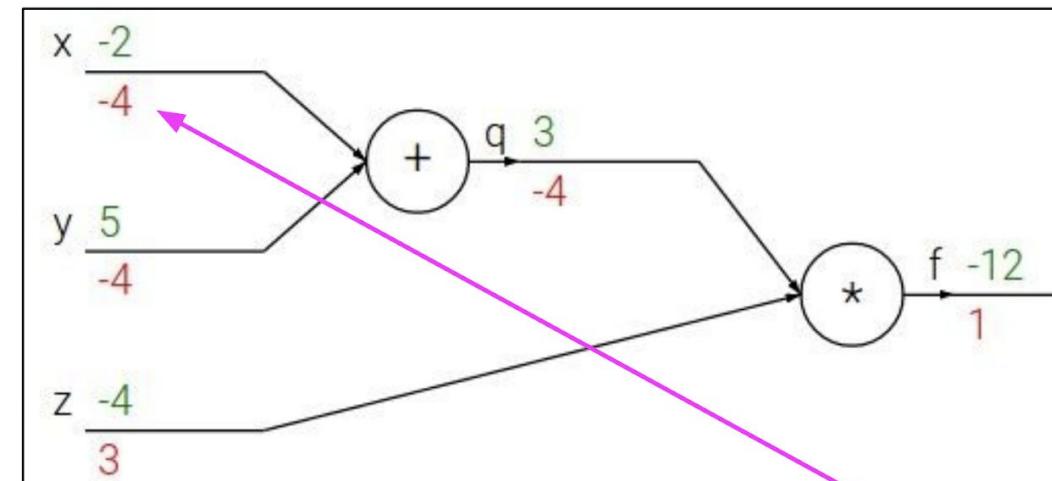
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Chain rule:

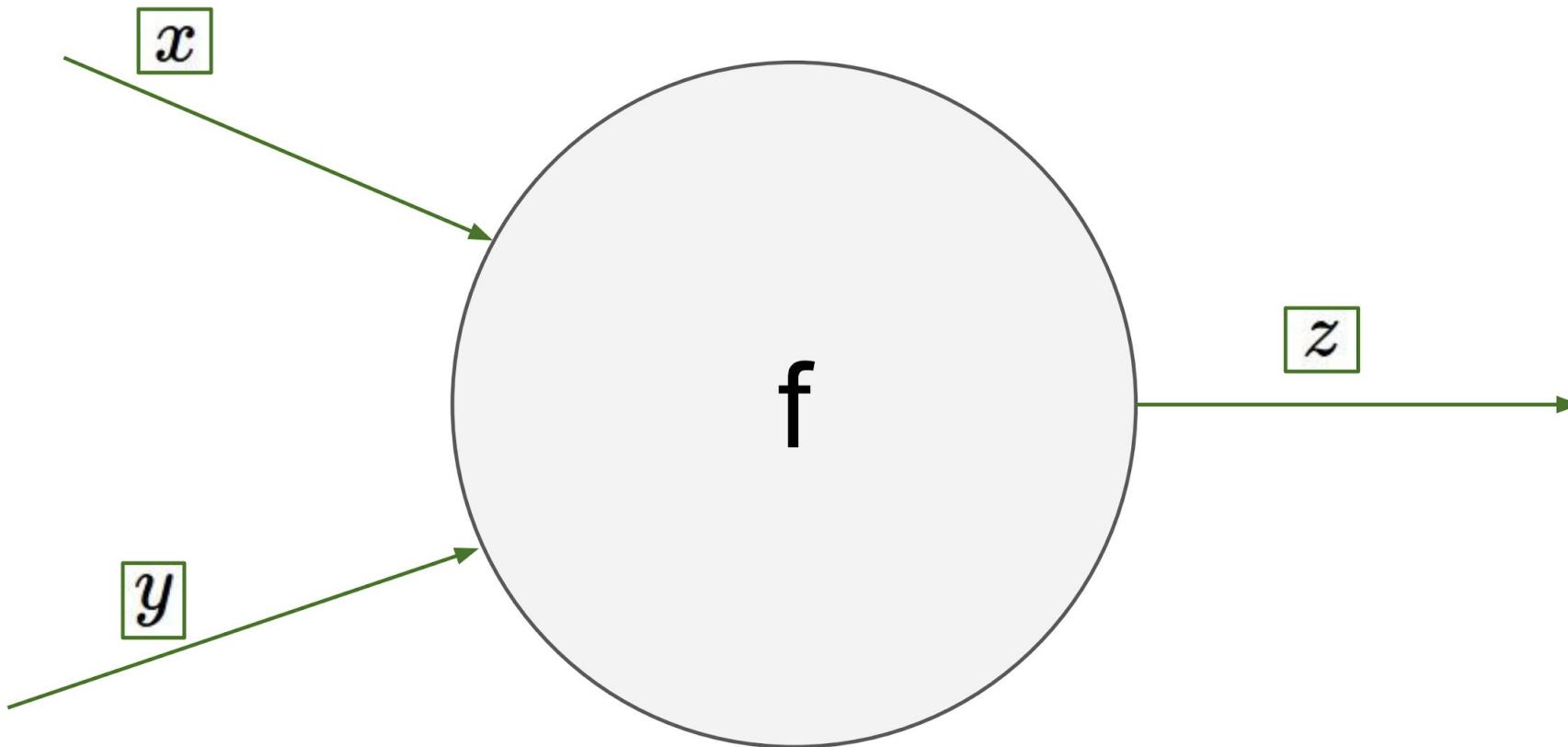
$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

Upstream
gradient

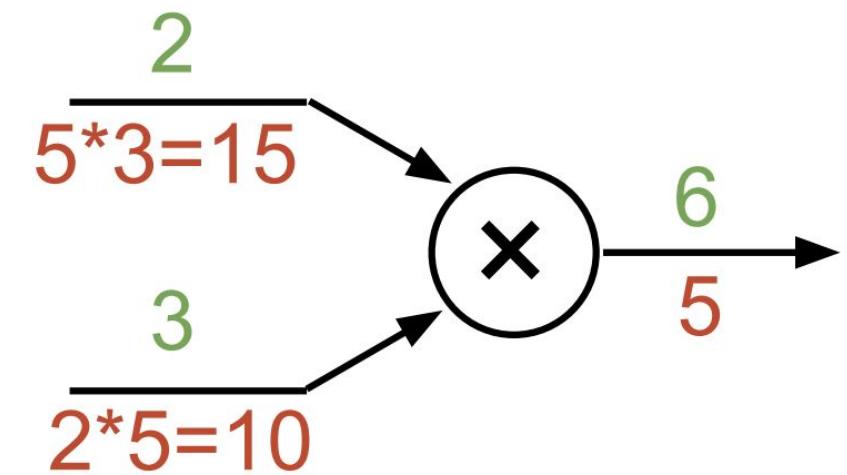
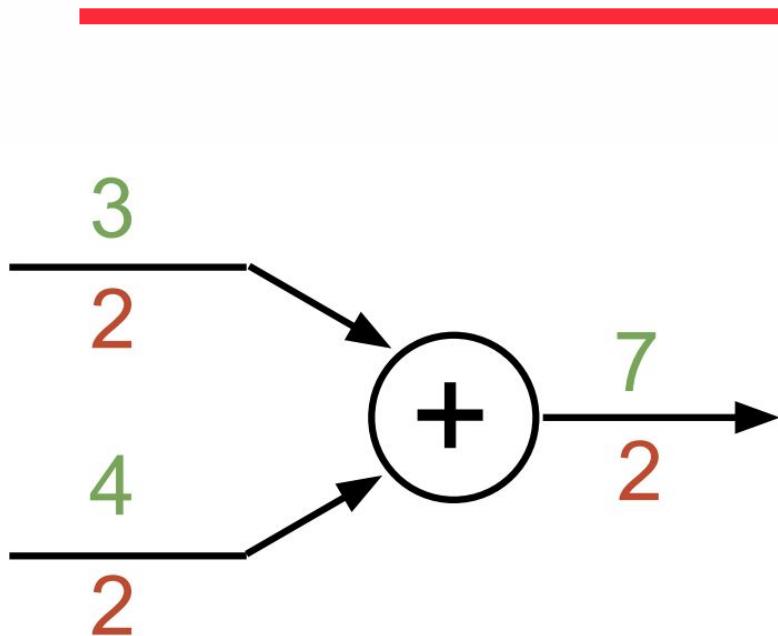
Local
gradient

$$\frac{\partial f}{\partial x}$$

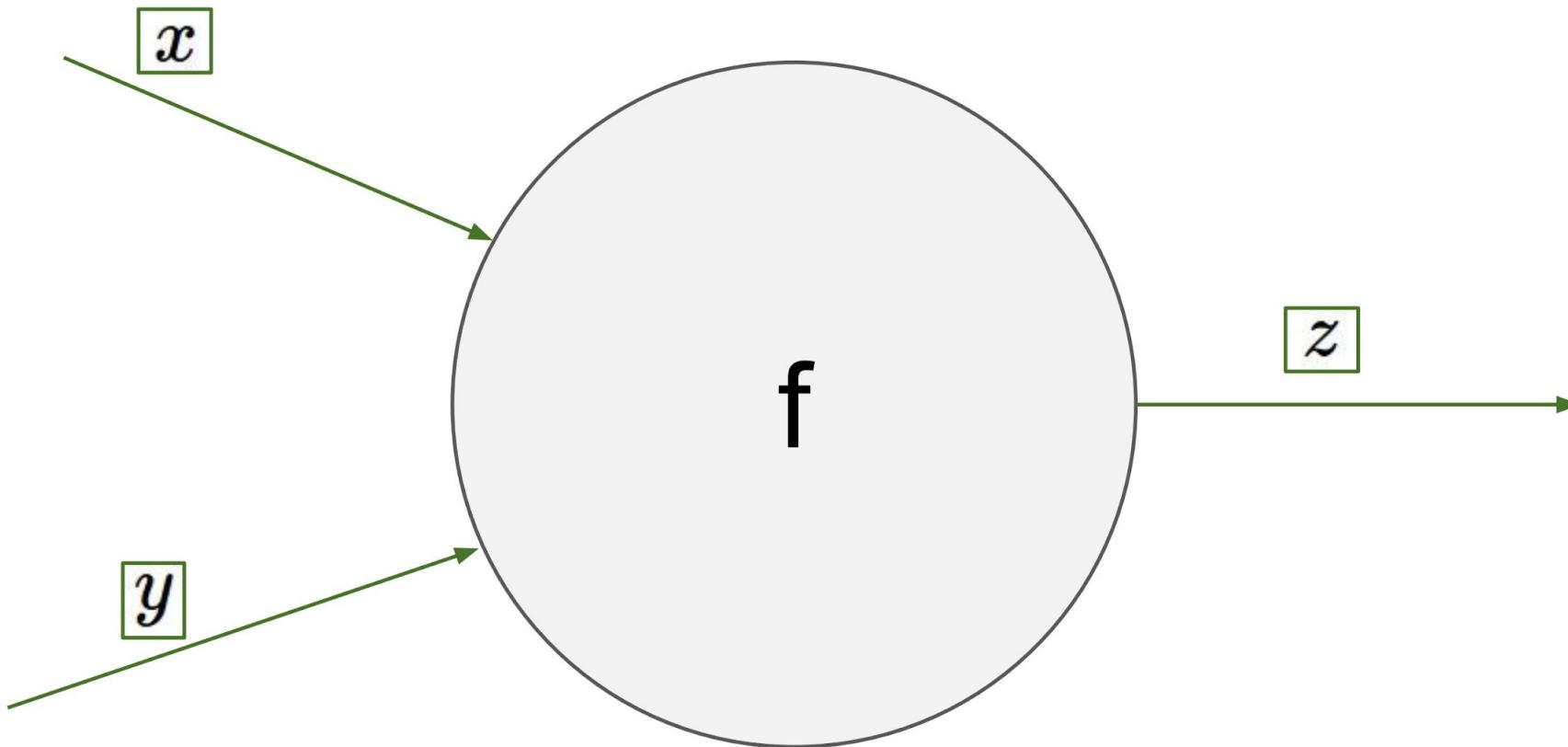
Кэширование градиентов при форвард пассе



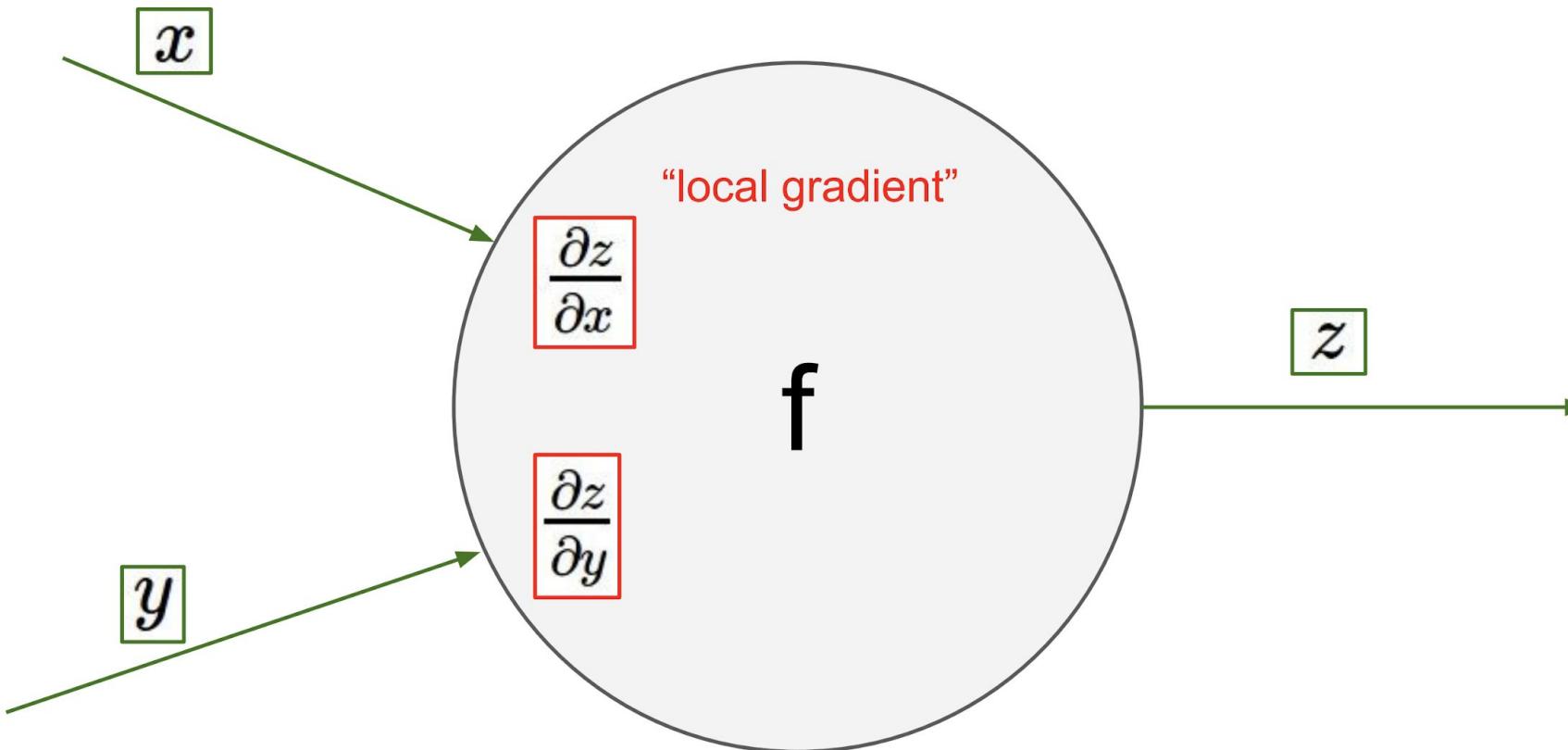
Кэширование градиентов при форвард пассе



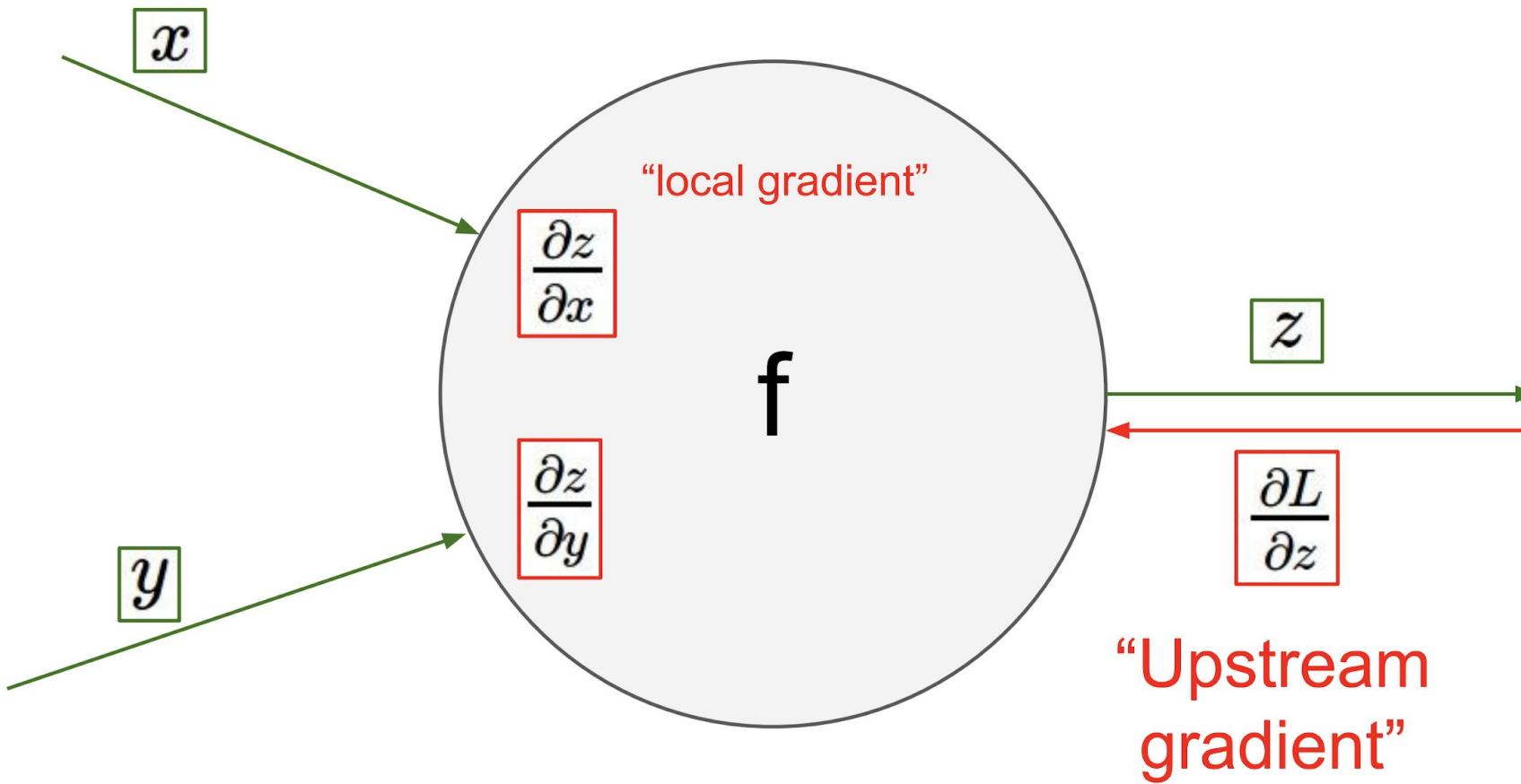
Кэширование градиентов при форвард пассе



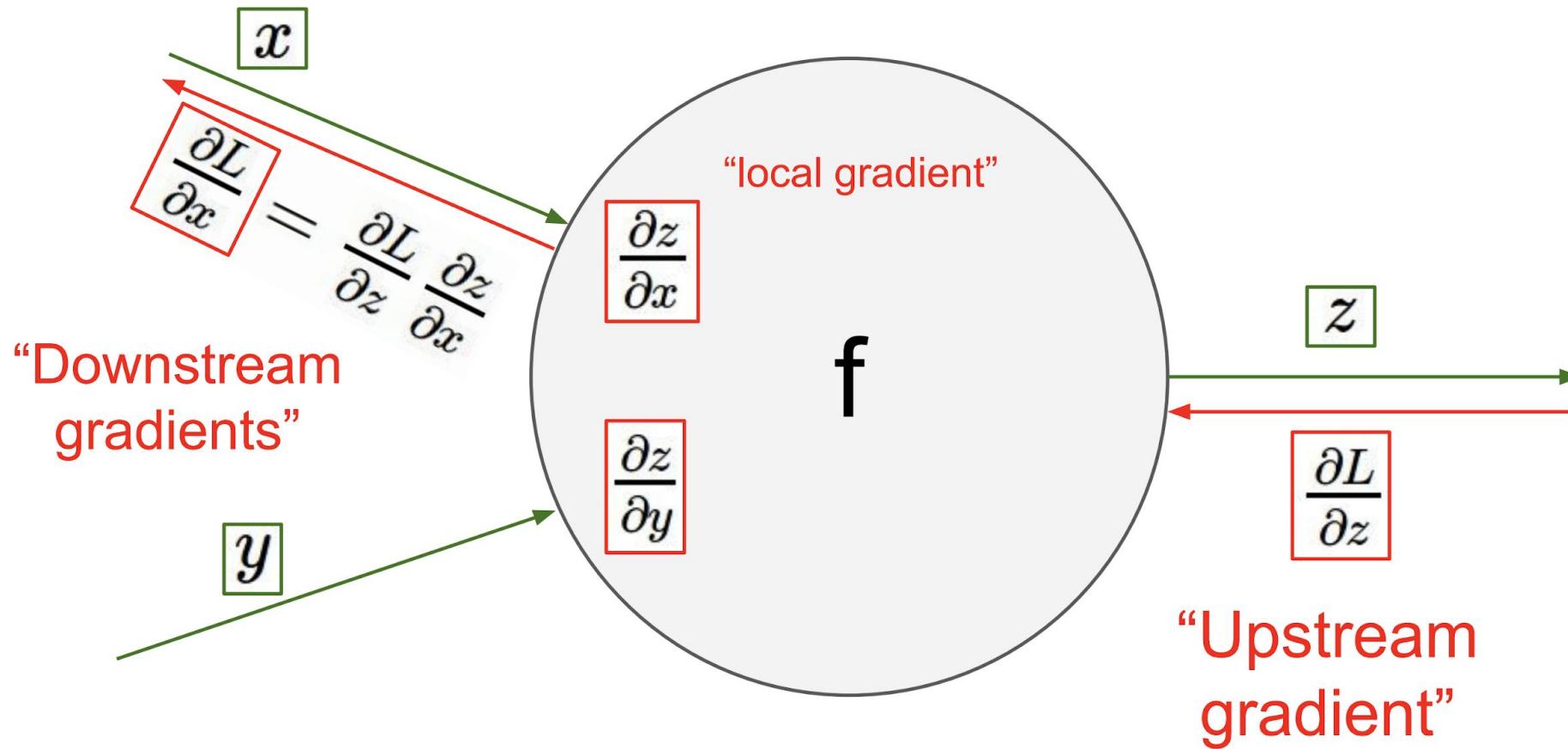
Кэширование градиентов при форвард пассе



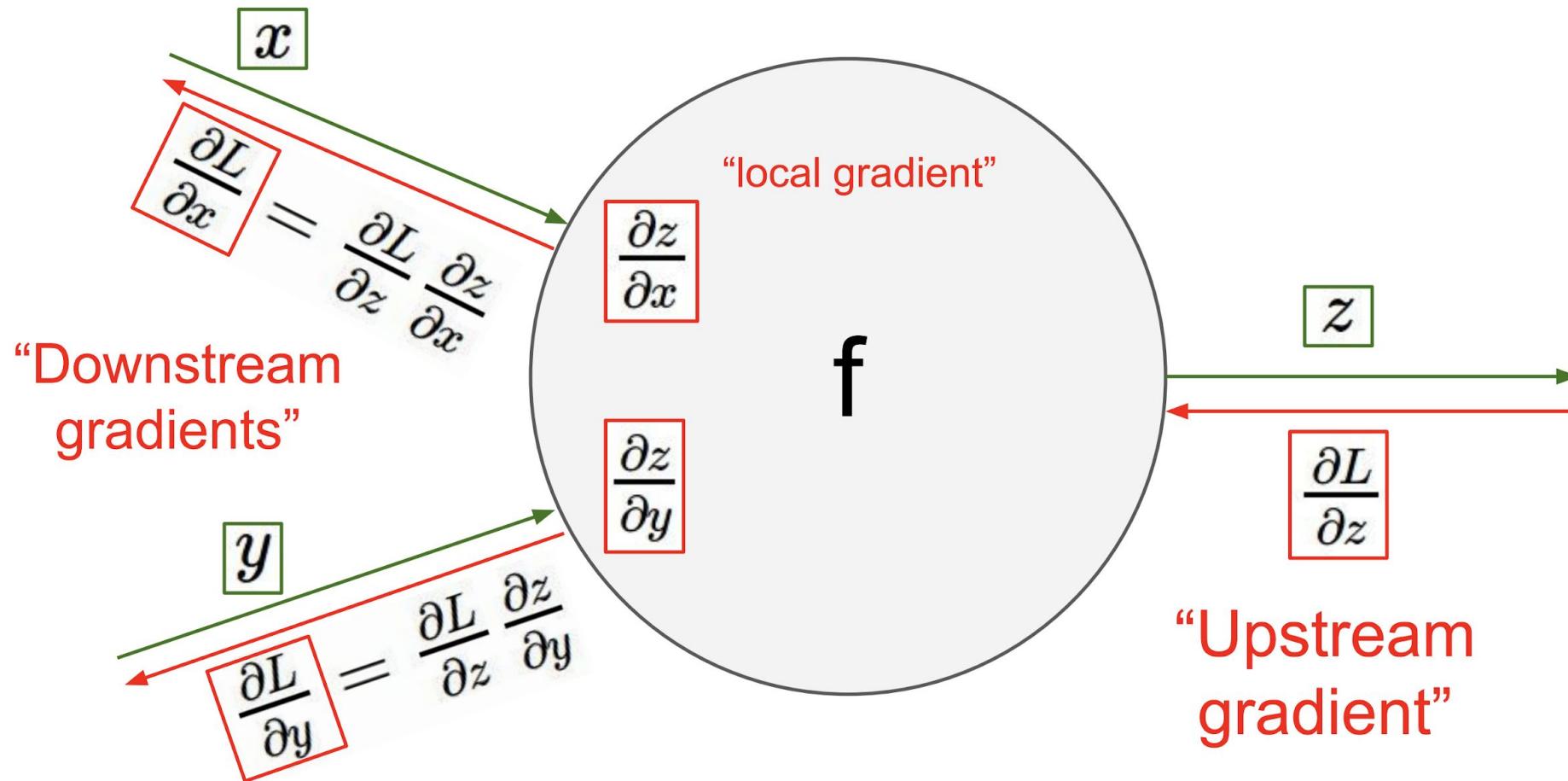
Кэширование градиентов при форвард пассе



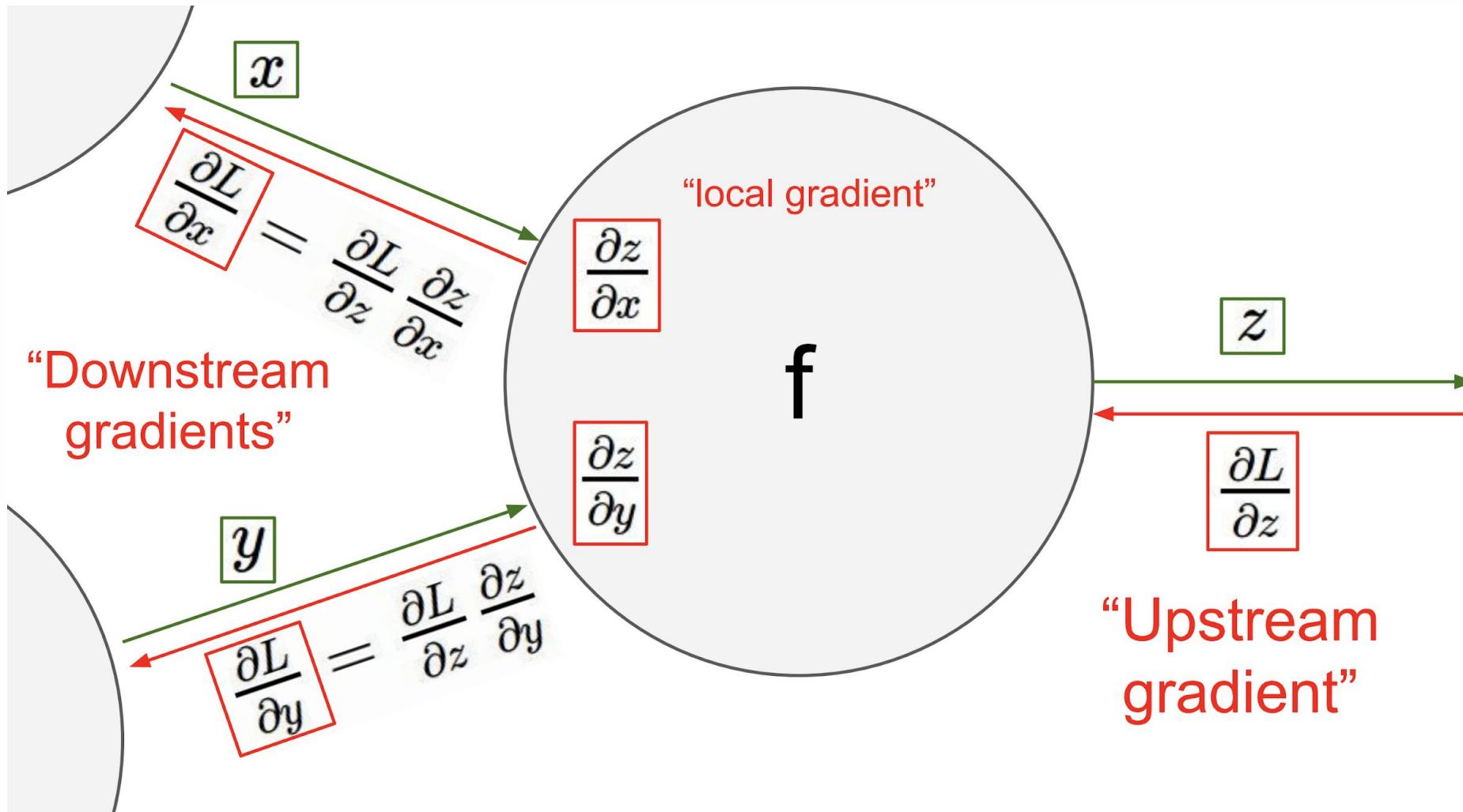
Кэширование градиентов при форвард пассе



Кэширование градиентов при форвард пассе



Кэширование градиентов при форвард пассе





Forward pass. Входной вектор: h

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$



Forward pass: step 1

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$



Forward pass: step 1. Beca

$$\theta = \boxed{U} h + \boxed{b_2}$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$



Forward pass: логирование градиентов

$$\theta = \boxed{U} h + \boxed{b_2}$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

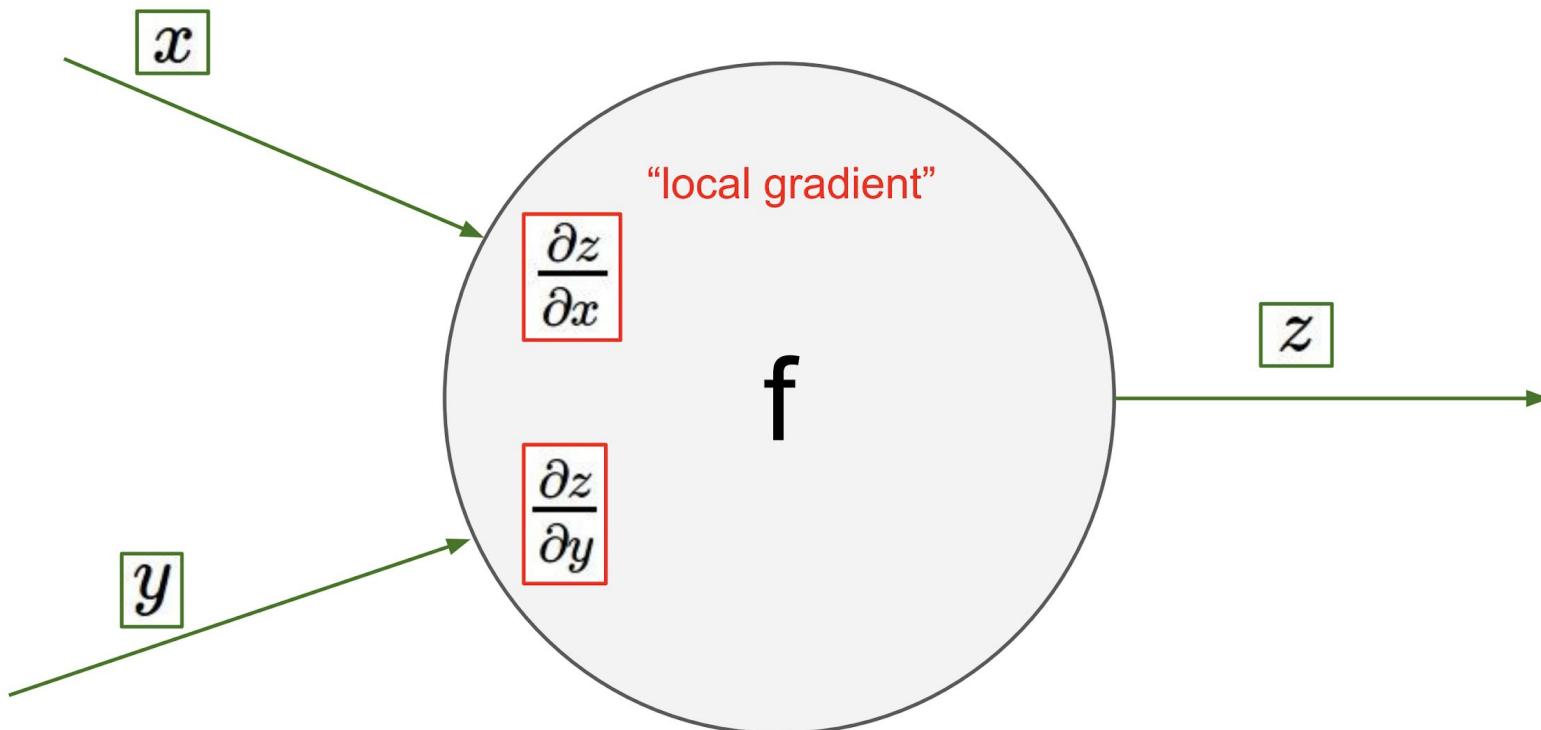
$$\boxed{\frac{\partial \theta}{\partial U}}$$

$$\boxed{\frac{\partial \theta}{\partial b_2}}$$

локальные градиенты

Forward pass: логирование градиентов

$$\theta = \boxed{U} h + \boxed{b_2}$$



$$\boxed{\frac{\partial \theta}{\partial U}}$$

$$\boxed{\frac{\partial \theta}{\partial b_2}}$$

локальные градиенты



Forward pass: step 2

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$



Forward pass: логирование градиентов

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

$$\frac{\partial \hat{y}}{\partial \theta}$$

локальные градиенты



Forward pass: step 3

$$\theta = \mathbf{U}h + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

Forward pass: логирование градиентов

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

$$\frac{\partial J}{\partial \hat{y}}$$

локальные градиенты

Backward pass

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

$$\frac{\partial J}{\partial \hat{y}}$$

локальные градиенты



Backward pass

$$\theta = \mathbf{U}h + \mathbf{b}_2$$

$$\hat{\mathbf{y}} = \text{softmax}(\theta)$$

$$J = CE(\mathbf{y}, \hat{\mathbf{y}})$$

$$\frac{\partial J}{\partial \mathbf{U}} =$$

$$\frac{\partial J}{\partial \mathbf{b}_2} =$$

Backward pass

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

$$\frac{\partial J}{\partial U} = \boxed{\frac{\partial J}{\partial \hat{y}}}$$
$$\frac{\partial J}{\partial b_2} = \boxed{\frac{\partial J}{\partial \hat{y}}}$$

Backward pass

$$\theta = Uh + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

$$\frac{\partial J}{\partial U} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta}$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta}$$

Backward pass

$$\theta = \mathbf{U}h + b_2$$

$$\hat{y} = \text{softmax}(\theta)$$

$$J = CE(y, \hat{y})$$

$$\frac{\partial J}{\partial \mathbf{U}} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \boxed{\frac{\partial \theta}{\partial \mathbf{U}}}$$

$$\frac{\partial J}{\partial b_2} = \frac{\partial J}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial \theta} \boxed{\frac{\partial \theta}{\partial b_2}}$$



Backward pass. Размерности

$$\theta = \mathbf{U}h + \mathbf{b}_2$$

$$\hat{\mathbf{y}} = \text{softmax}(\theta)$$

$$J = CE(\mathbf{y}, \hat{\mathbf{y}})$$

$$\frac{\partial J}{\partial \mathbf{U}} = \frac{\partial J}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{U}}$$

$$\frac{\partial J}{\partial \mathbf{b}_2} = \frac{\partial J}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{b}_2}$$



Backward pass. Размерности

$$\theta = \mathbf{U}h + \mathbf{b}_2$$

$$\hat{\mathbf{y}} = \text{softmax}(\theta)$$

$$J = CE(\mathbf{y}, \hat{\mathbf{y}})$$

$$\mathbf{b}_2 \in \mathbb{R}^{N_c \times 1}$$

$$\mathbf{U} \in \mathbb{R}^{N_c \times D_h}$$

$$\frac{\partial J}{\partial \mathbf{U}} = \frac{\partial J}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{U}}$$

$$\frac{\partial J}{\partial \mathbf{b}_2} = \frac{\partial J}{\partial \hat{\mathbf{y}}} \frac{\partial \hat{\mathbf{y}}}{\partial \theta} \frac{\partial \theta}{\partial \mathbf{b}_2}$$

Backward pass. Одномерный случай

Backpropagation: a simple example

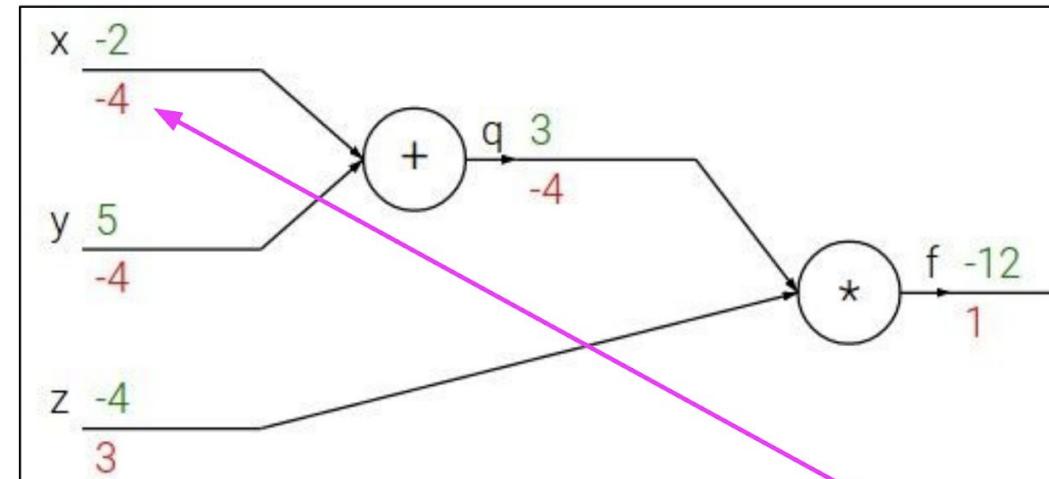
$$f(x, y, z) = (x + y)z$$

e.g. $x = -2$, $y = 5$, $z = -4$

$$q = x + y \quad \frac{\partial q}{\partial x} = 1, \frac{\partial q}{\partial y} = 1$$

$$f = qz \quad \frac{\partial f}{\partial q} = z, \frac{\partial f}{\partial z} = q$$

Want: $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$



Chain rule:

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial q} \frac{\partial q}{\partial x}$$

Upstream gradient Local gradient

$$\frac{\partial f}{\partial x}$$



Backward pass. Двумерный случай



Backward pass. Двумерный случай

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$$



Backward pass. Двумерный случай

$$\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\mathbf{f}(\mathbf{x}) = [f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)]$$

Backward pass. Двумерный случай

$$f : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$f(\mathbf{x}) = [f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)]$$

$$f(\mathbf{x}) = \begin{bmatrix} w_{01} & w_{02} & w_{03} \\ w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Backward pass. Двумерный случай

$$\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\mathbf{f}(\mathbf{x}) = [f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)]$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} w_{01} & w_{02} & w_{03} \\ w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Якобиан

$$\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\mathbf{f}(\mathbf{x}) = [f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)]$$

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} w_{01} & w_{02} & w_{03} \\ w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \end{bmatrix}$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Якобиан

$$\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$\mathbf{f}(\mathbf{x}) = [f_1(x_1, \dots, x_n), f_2(x_1, \dots, x_n), \dots, f_m(x_1, \dots, x_n)]$$

$$\mathbf{f}(\mathbf{x}) = \begin{bmatrix} w_{01} & w_{02} & w_{03} \\ w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} \times \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \end{bmatrix}$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$



Backward pass. Двумерный случай. Chain rule

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$



Backward pass. Двумерный случай. Chain rule

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$

$$\mathbf{g}(y) = [g_1(y_1, y_2), g_2(y_1, y_2)]$$

Backward pass. Двумерный случай. Chain rule

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$

$$\mathbf{g}(y) = [g_1(y_1, y_2), g_2(y_1, y_2)]$$

$$\mathbf{g}(x) = [g_1(f_1(x), f_2(x)), g_2(f_1(x), f_2(x))]$$

$$h(x) = \underbrace{(5 - 6x)}_{\text{outer}}^{\text{inner}}{}^5$$

$$g(x) = 5 - 6x \quad \text{inner function}$$

$$f(x) = x^5 \quad \text{outer function}$$

<http://web.stanford.edu/class/cs224n/readings/gradient-notes.pdf>



Backward pass. Двумерный случай. Chain rule

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$

$$\mathbf{g}(\mathbf{y}) = [g_1(y_1, y_2), g_2(y_1, y_2)]$$

$$\mathbf{g}(x) = [g_1(f_1(x), f_2(x)), g_2(f_1(x), f_2(x))]$$



Якобиан

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$

$$\mathbf{g}(y) = [g_1(y_1, y_2), g_2(y_1, y_2)]$$

$$\mathbf{g}(x) = [g_1(f_1(x), f_2(x)), g_2(f_1(x), f_2(x))]$$

$$\frac{\partial \mathbf{g}}{\partial x} =$$



Якобиан

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$

$$\mathbf{g}(y) = [g_1(y_1, y_2), g_2(y_1, y_2)]$$

$$\mathbf{g}(x) = [g_1(f_1(x), f_2(x)), g_2(f_1(x), f_2(x))]$$

$$\frac{\partial \mathbf{g}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} g_1(f_1(x), f_2(x)) \\ \frac{\partial}{\partial x} g_2(f_1(x), f_2(x)) \end{bmatrix}$$

Якобиан

$$\mathbf{f}(x) = [f_1(x), f_2(x)]$$

$$\mathbf{g}(y) = [g_1(y_1, y_2), g_2(y_1, y_2)]$$

$$\mathbf{g}(x) = [g_1(f_1(x), f_2(x)), g_2(f_1(x), f_2(x))]$$

$$\frac{\partial \mathbf{g}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} g_1(f_1(x), f_2(x)) \\ \frac{\partial}{\partial x} g_2(f_1(x), f_2(x)) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_1}{\partial f_2} \frac{\partial f_2}{\partial x} \\ \frac{\partial g_2}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_2}{\partial f_2} \frac{\partial f_2}{\partial x} \end{bmatrix}$$



Повторение: Якобиан

$$\frac{\partial \mathbf{g}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} g_1(f_1(x), f_2(x)) \\ \frac{\partial}{\partial x} g_2(f_1(x), f_2(x)) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_1}{\partial f_2} \frac{\partial f_2}{\partial x} \\ \frac{\partial g_2}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_2}{\partial f_2} \frac{\partial f_2}{\partial x} \end{bmatrix}$$

Якобиан

$$\frac{\partial \mathbf{g}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} g_1(f_1(x), f_2(x)) \\ \frac{\partial}{\partial x} g_2(f_1(x), f_2(x)) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_1}{\partial f_2} \frac{\partial f_2}{\partial x} \\ \frac{\partial g_2}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_2}{\partial f_2} \frac{\partial f_2}{\partial x} \end{bmatrix}$$

$$\frac{\partial \mathbf{g}}{\partial x} = \frac{\partial \mathbf{g}}{\partial \mathbf{f}} \frac{\partial \mathbf{f}}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} & \frac{\partial g_1}{\partial f_2} \\ \frac{\partial g_2}{\partial f_1} & \frac{\partial g_2}{\partial f_2} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \end{bmatrix}$$

Якобиан

$$\frac{\partial \mathbf{g}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} g_1(f_1(x), f_2(x)) \\ \frac{\partial}{\partial x} g_2(f_1(x), f_2(x)) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_1}{\partial f_2} \frac{\partial f_2}{\partial x} \\ \frac{\partial g_2}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_2}{\partial f_2} \frac{\partial f_2}{\partial x} \end{bmatrix}$$

$$\frac{\partial \mathbf{g}}{\partial x} = \frac{\partial \mathbf{g}}{\partial \mathbf{f}} \frac{\partial \mathbf{f}}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} & \frac{\partial g_1}{\partial f_2} \\ \frac{\partial g_2}{\partial f_1} & \frac{\partial g_2}{\partial f_2} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \end{bmatrix}$$

$$\frac{d}{dx} [f(g(x))] = f'(g(x)) \cdot g'(x)$$

Якобиан

$$\frac{\partial \mathbf{g}}{\partial x} = \begin{bmatrix} \frac{\partial}{\partial x} g_1(f_1(x), f_2(x)) \\ \frac{\partial}{\partial x} g_2(f_1(x), f_2(x)) \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_1}{\partial f_2} \frac{\partial f_2}{\partial x} \\ \frac{\partial g_2}{\partial f_1} \frac{\partial f_1}{\partial x} + \frac{\partial g_2}{\partial f_2} \frac{\partial f_2}{\partial x} \end{bmatrix}$$

$$\frac{\partial \mathbf{g}}{\partial x} = \frac{\partial \mathbf{g}}{\partial \mathbf{f}} \frac{\partial \mathbf{f}}{\partial x} = \begin{bmatrix} \frac{\partial g_1}{\partial f_1} & \frac{\partial g_1}{\partial f_2} \\ \frac{\partial g_2}{\partial f_1} & \frac{\partial g_2}{\partial f_2} \end{bmatrix} \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \frac{\partial f_2}{\partial x} \end{bmatrix}$$

$$\frac{d}{dx} [f(g(x))] = f'(g(x)) \cdot g'(x)$$

$$\begin{aligned} & \frac{d}{dx} [f(g(x))] \\ &= f'(g(x)) \cdot g'(x) \\ &= 5(5 - 6x)^4 \cdot -6 \\ &= -30(5 - 6x)^4 \end{aligned}$$



Backprop. Пример расчета градиента

$$\mathbf{z} = \mathbf{W}\mathbf{x}$$



Backprop. Пример расчета градиента

$$\frac{\partial z}{\partial x} ?$$

$$z = \mathbf{Wx}$$



Backprop. Пример расчета градиента

$$\mathbf{z} = \mathbf{W}\mathbf{x} \quad z_i = \sum_{k=1}^m W_{ik}x_k$$

$\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$?



Backprop. Пример расчета градиента

$\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$?

$$\mathbf{z} = \mathbf{W}\mathbf{x} \quad z_i = \sum_{k=1}^m W_{ik}x_k$$

$$(\frac{\partial \mathbf{z}}{\partial \mathbf{x}})_{ij} =$$



Backprop. Пример расчета градиента

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} ?$$

$$\mathbf{z} = \mathbf{W}\mathbf{x} \quad z_i = \sum_{k=1}^m W_{ik}x_k$$

$$(\frac{\partial \mathbf{z}}{\partial \mathbf{x}})_{ij} = \frac{\partial z_i}{\partial x_j} =$$

 $\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$?

Backprop. Пример расчета градиента

$$\mathbf{z} = \mathbf{W}\mathbf{x} \quad z_i = \sum_{k=1}^m W_{ik}x_k$$

$$(\frac{\partial \mathbf{z}}{\partial \mathbf{x}})_{ij} = \frac{\partial z_i}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{k=1}^m W_{ik}x_k =$$



Backprop. Пример расчета градиента

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} ?$$

$$\mathbf{z} = \mathbf{Wx} \quad z_i = \sum_{k=1}^m W_{ik} x_k$$

$$(\frac{\partial \mathbf{z}}{\partial \mathbf{x}})_{ij} = \frac{\partial z_i}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{k=1}^m W_{ik} x_k = \sum_{k=1}^m W_{ik} \frac{\partial}{\partial x_j} x_k =$$



Backprop. Пример расчета градиента

$\frac{\partial \mathbf{z}}{\partial \mathbf{x}}$?

$$\mathbf{z} = \mathbf{Wx} \quad z_i = \sum_{k=1}^m W_{ik} x_k$$

$$(\frac{\partial \mathbf{z}}{\partial \mathbf{x}})_{ij} = \frac{\partial z_i}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{k=1}^m W_{ik} x_k = \sum_{k=1}^m W_{ik} \frac{\partial}{\partial x_j} x_k = W_{ij}$$



Backprop. Пример расчета градиента

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} ?$$

$$\mathbf{z} = \mathbf{Wx} \quad z_i = \sum_{k=1}^m W_{ik} x_k$$

$$(\frac{\partial \mathbf{z}}{\partial \mathbf{x}})_{ij} = \frac{\partial z_i}{\partial x_j} = \frac{\partial}{\partial x_j} \sum_{k=1}^m W_{ik} x_k = \sum_{k=1}^m W_{ik} \frac{\partial}{\partial x_j} x_k = W_{ij}$$

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \mathbf{W}$$



Резюме

- Полносвязные сети (Fully-connected neural networks) - сеть из N блоков, где каждый блок состоит из линейной операции (умножение на матрицу) и нелинейной функции активации
- Примеры нелинейных функций: tanh, sigmoid, softmax, relu
- Backpropagation - рекурсивный алгоритм подсчета градиентов в сети, использующий chain rule
- forward pass - подсчет выходов из каждого слоя сети
- backward pass - подсчет градиентов и обновление весов сети



В следующих сериях

- Лекция о разных алгоритмах оптимизации и регуляризации сетей
- Семинар: введение в Pytorch, полносвязная сеть на Pytorch



Полезные ссылки

- Лекция по backpropagation, CS231n, CV with Deep Learning:
<https://www.youtube.com/watch?v=i94OvYb6noo>
- Лекция по backpropagation, CS224n, NLP with Deep Learning:
<https://www.youtube.com/watch?v=yLYHDSv-288>
- Затеханный семинар по backprop из Стэнфорда:
<http://web.stanford.edu/class/cs224n/readings/gradient-notes.pdf>