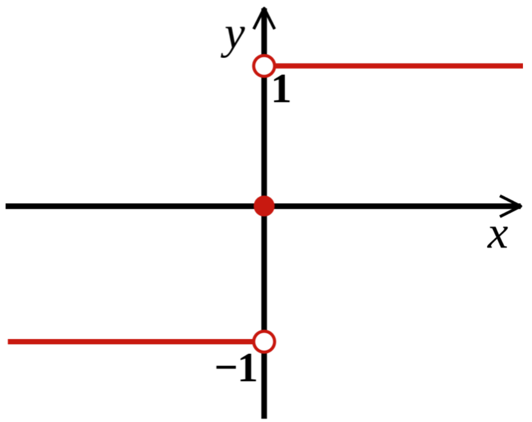
MML minor #3

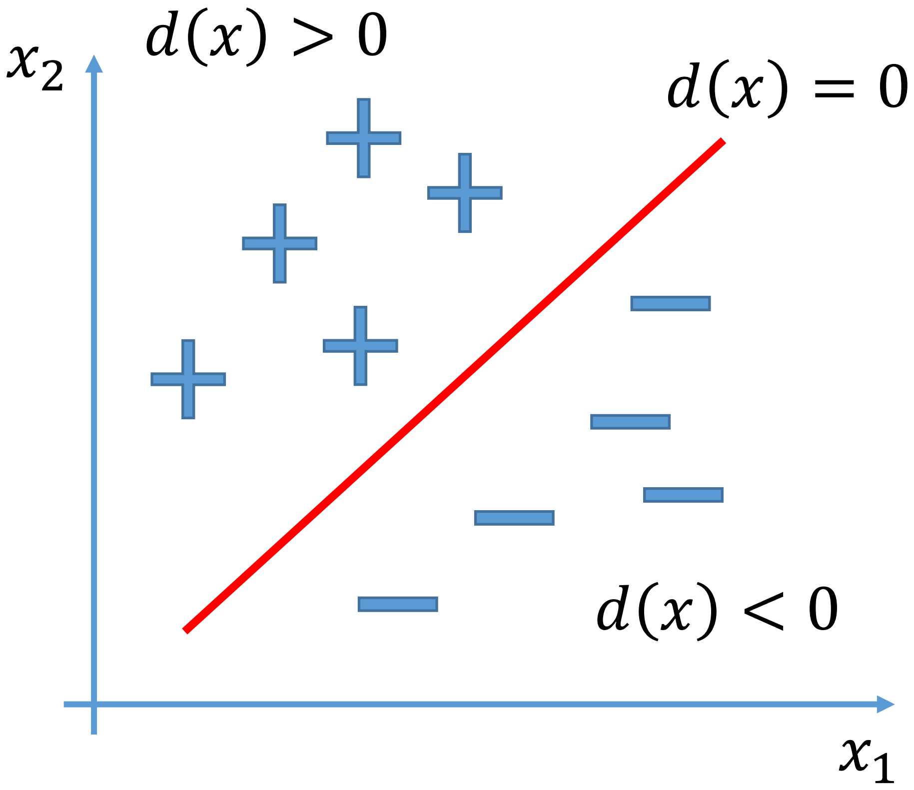
Нейронные сети

Линейная классификация

- Признаки: $x = (x_1, x_2)$
- Целевая переменная: $y \in \{+1, -1\}$
- Функция принятия решения: $d(x) = \boldsymbol{w_0} + \boldsymbol{w_1} x_1 + \boldsymbol{w_2} x_2$
- Алгоритм: $a(x) = \mathrm{sign}(d(x))$

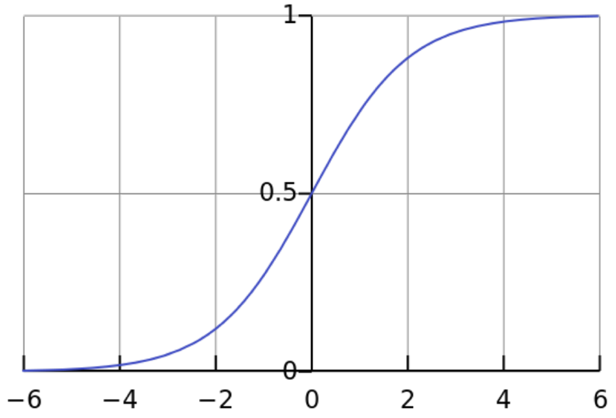


$\mathrm{sign}(x)$

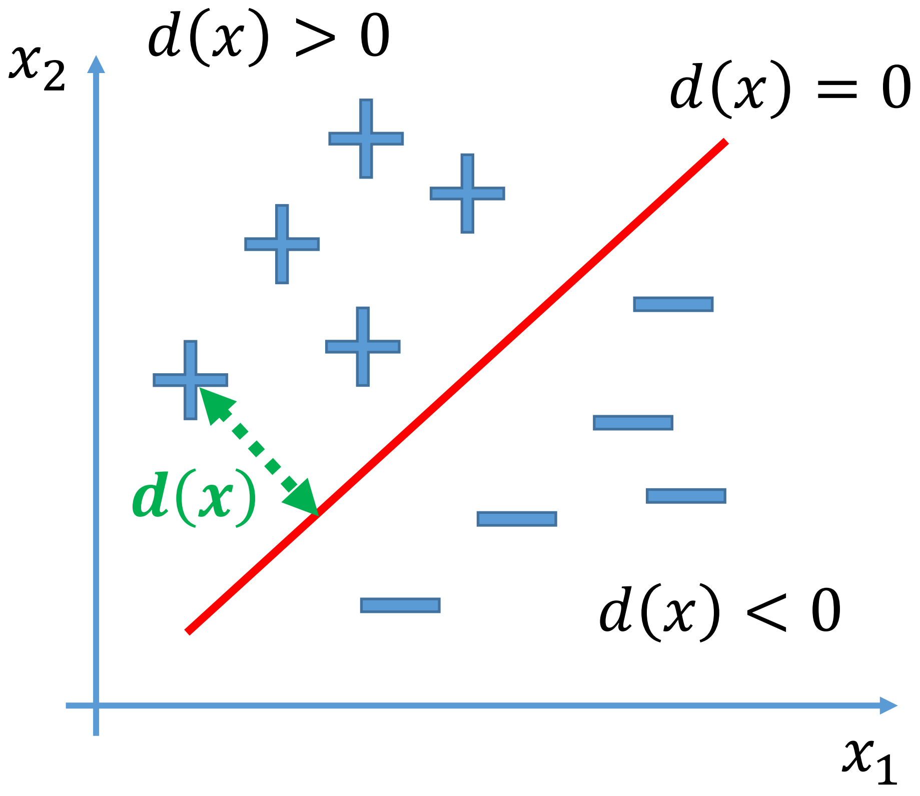


Логистическая регрессия

- Предсказывает вероятность положительного класса (+1)
- Функция принятия решения: $d(x) = \boldsymbol{w_0} + \boldsymbol{w_1} x_1 + \boldsymbol{w_2} x_2$
- Алгоритм: $a(x) = \sigma\big(d(x)\big)$

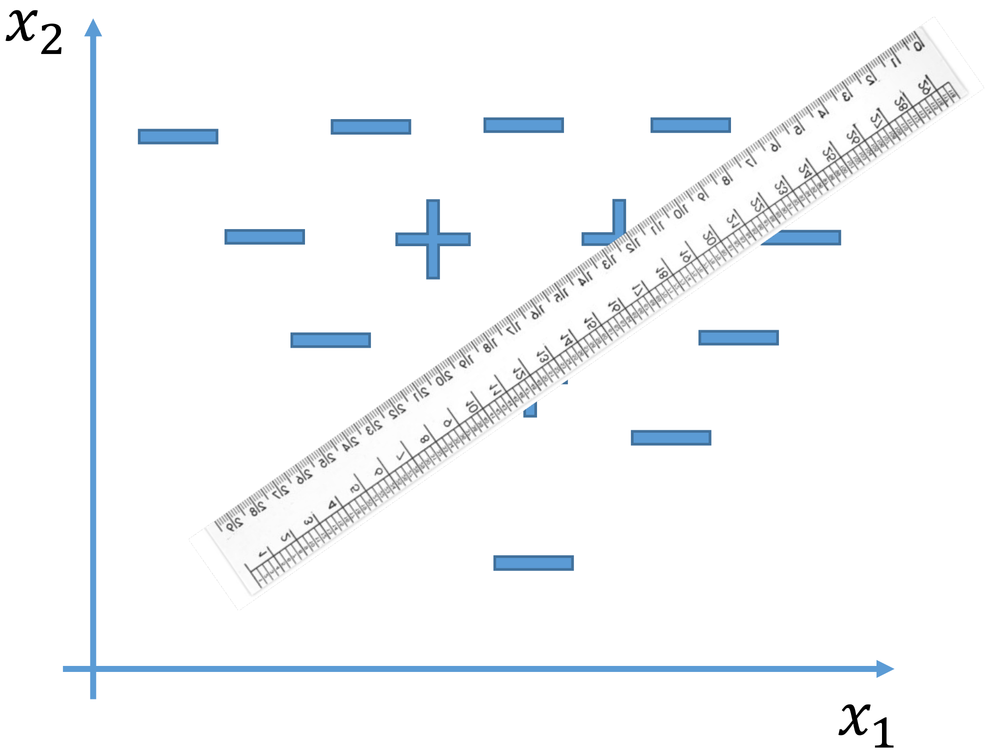


$$\sigma(x) = \frac{1}{1 + e^{-x}}$$



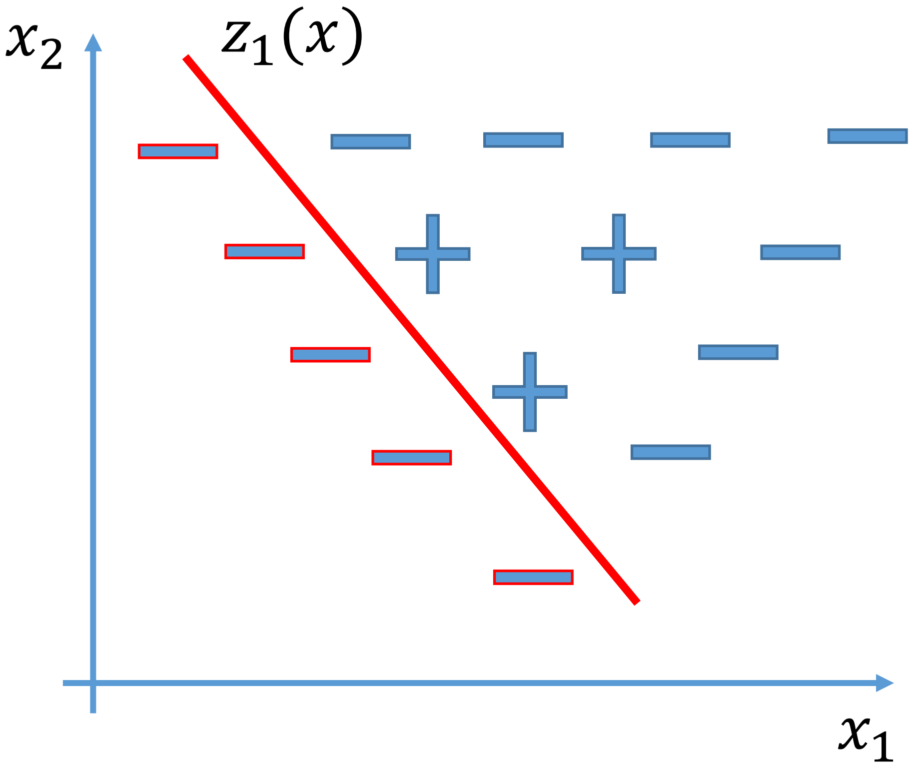
А что же делать тут?

- Признаки: $x = (x_1, x_2)$
- Целевой признак: $y \in \{+1, -1\}$



Решим подзадачу

- Признаки: $x = (x_1, x_2)$
- Целевой признак: $y \in \{+1, -1\}$

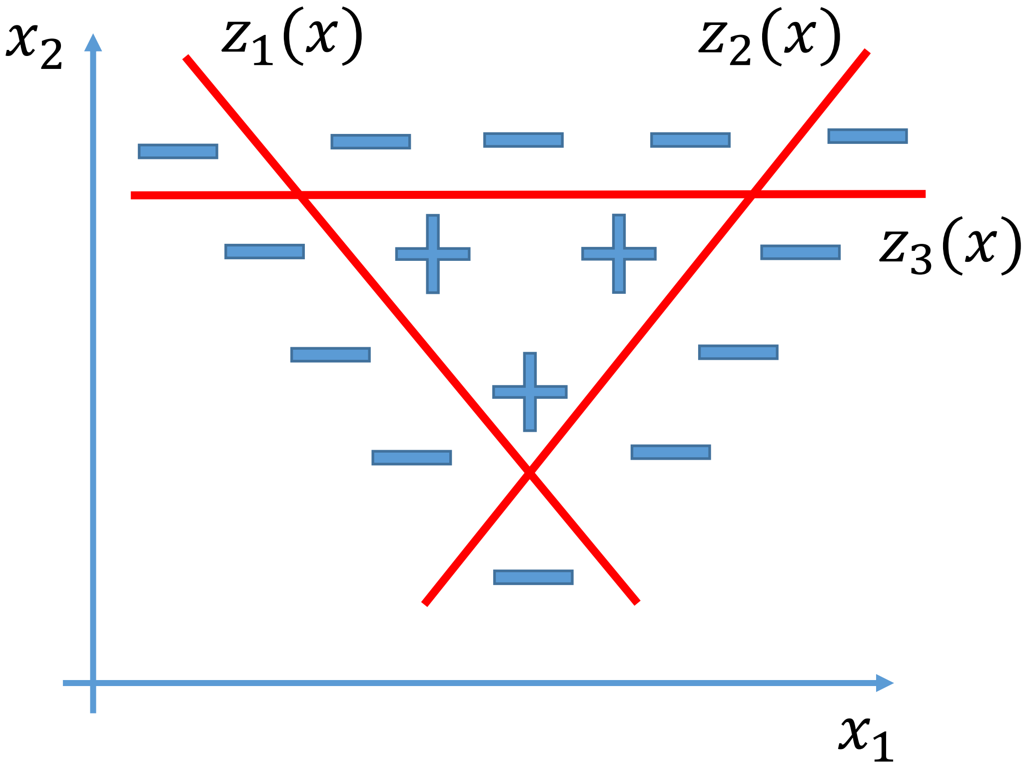


Отделим минусы слева

$$z_1 = \sigma(\boldsymbol{w_{0,1}} + \boldsymbol{w_{1,1}} x_1 + \boldsymbol{w_{2,1}} x_2)$$

Своя линия для каждой подзадачи

- Признаки: $x = (x_1, x_2)$
- Целевой признак: $y \in \{+1, -1\}$

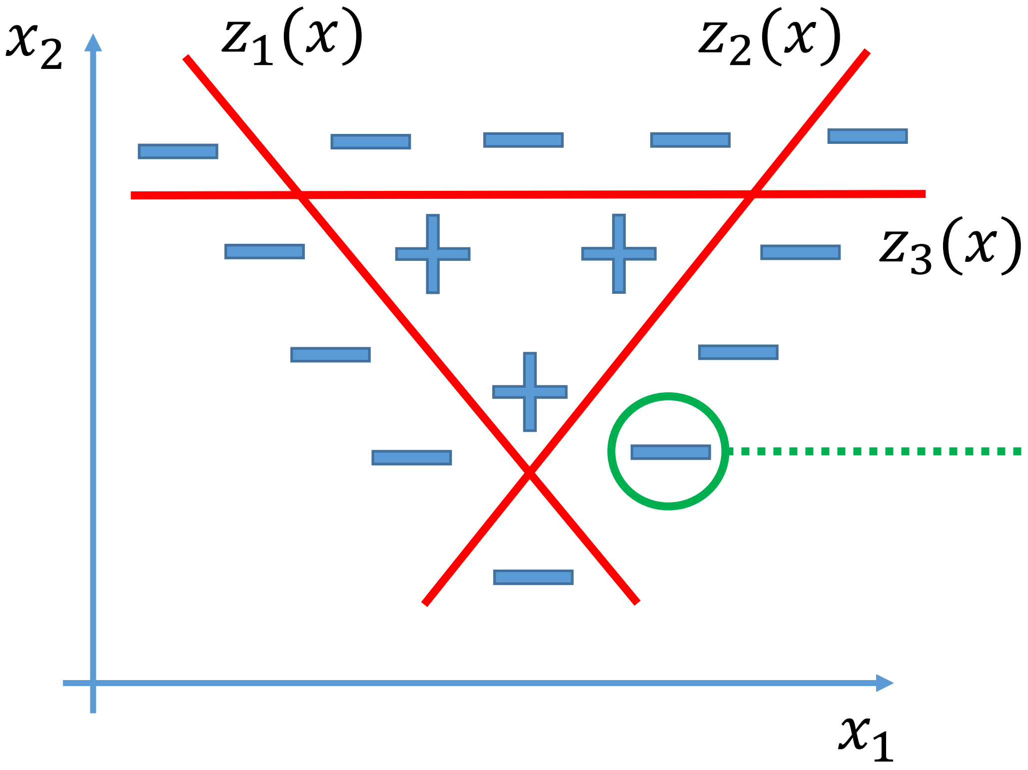


$$z_i = \sigma(\boldsymbol{w_{0,i}} + \boldsymbol{w_{1,i}} x_1 + \boldsymbol{w_{2,i}} x_2)$$

Допустим, мы нашли 3 таких линии…

Используем предсказания линий

- Признаки: $x = (x_1, x_2)$
- Целевой признак: $y \in \{+1, -1\}$



$$z_i = \sigma(\boldsymbol{w_{0,i}} + \boldsymbol{w_{1,i}} x_1 + \boldsymbol{w_{2,i}} x_2)$$

Новые признаки:

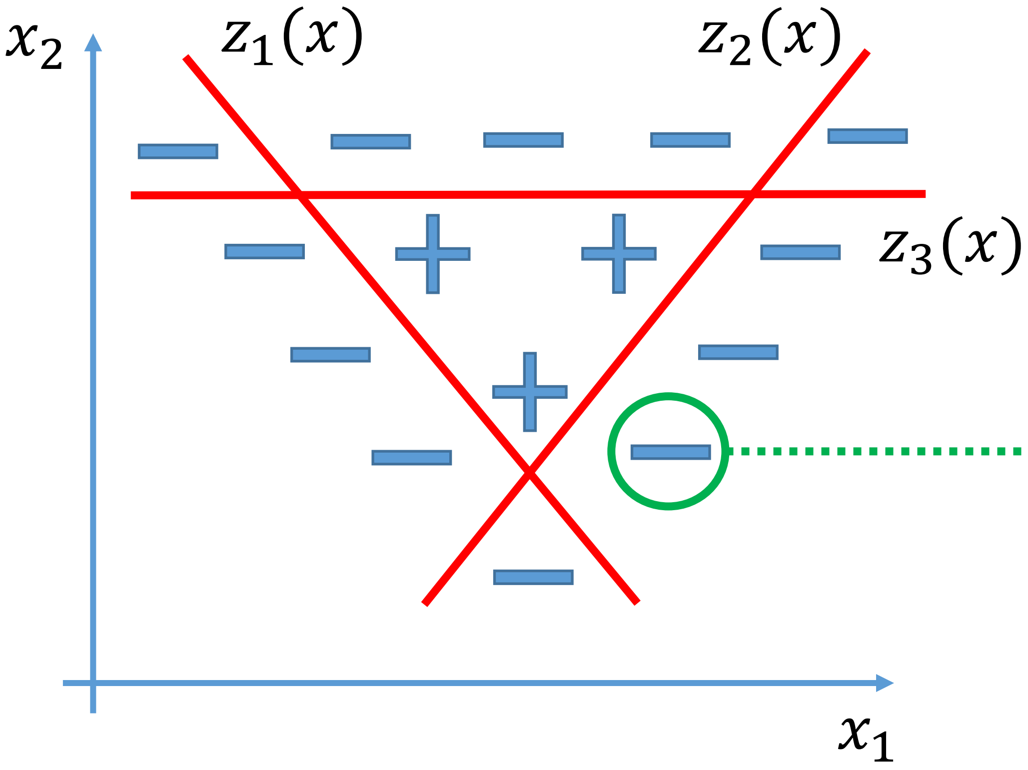
$z_1(x)$	$z_2(x)$	$z_3(x)$	y
0.6	**0.3**	**0.8**	**-1**
0.7	**0.7**	**0.7**	**+1**

$$(x_1, x_2) \to (z_1, z_2, z_3)$$

Что дальше?

Финальная модель

- Признаки: $x = (x_1, x_2)$
- Целевой признак: $y \in \{+1, -1\}$



$$z_i = \sigma(\boldsymbol{w_{0,i}} + \boldsymbol{w_{1,i}} x_1 + \boldsymbol{w_{2,i}} x_2)$$

Новые признаки:

$z_1(x)$	$z_2(x)$	$z_3(x)$	y
0.6	0.3	0.8	-1
0.7	0.7	0.7	+1

$$(x_1, x_2) \to (z_1, z_2, z_3)$$

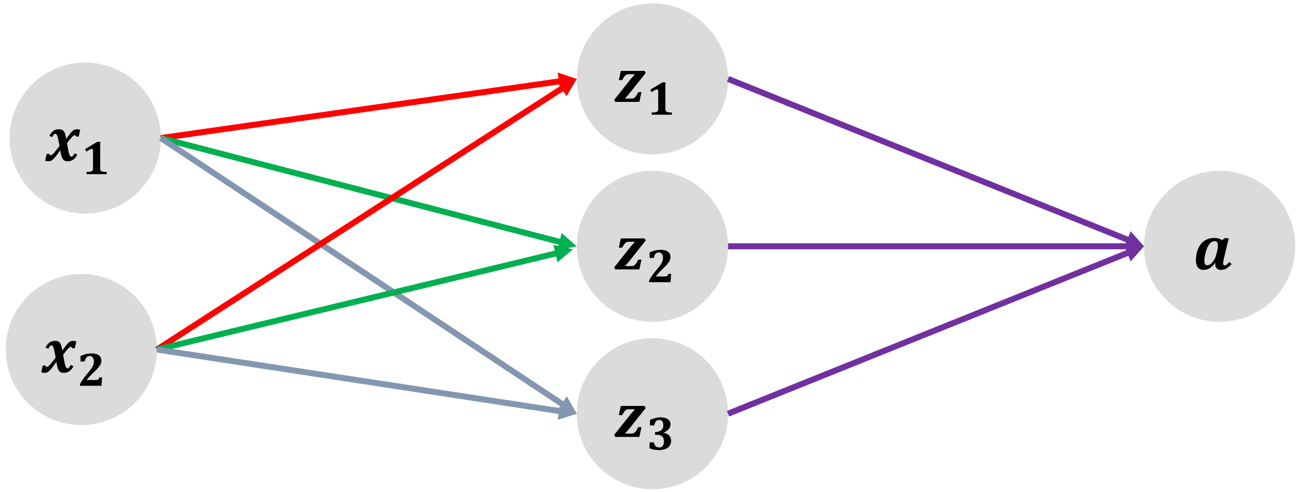
Строим финальную линейную модель:

$$a(x) = \sigma(\boldsymbol{w_0} + \boldsymbol{w_1} z_1(x) + \boldsymbol{w_2} z_2(x) + \boldsymbol{w_3} z_3(x))$$

Мы все еще не знаем как найти параметры линий

- Но понятно как все будет работать, если найти эти параметры:
  - $z_i = \sigma(\boldsymbol{w_{0,i}} + \boldsymbol{w_{1,i}} x_1 + \boldsymbol{w_{2,i}} x_2)$
  - $a(x) = \sigma(\boldsymbol{w_0} + \boldsymbol{w_1} z_1(x) + \boldsymbol{w_2} z_2(x) + \boldsymbol{w_3} z_3(x))$

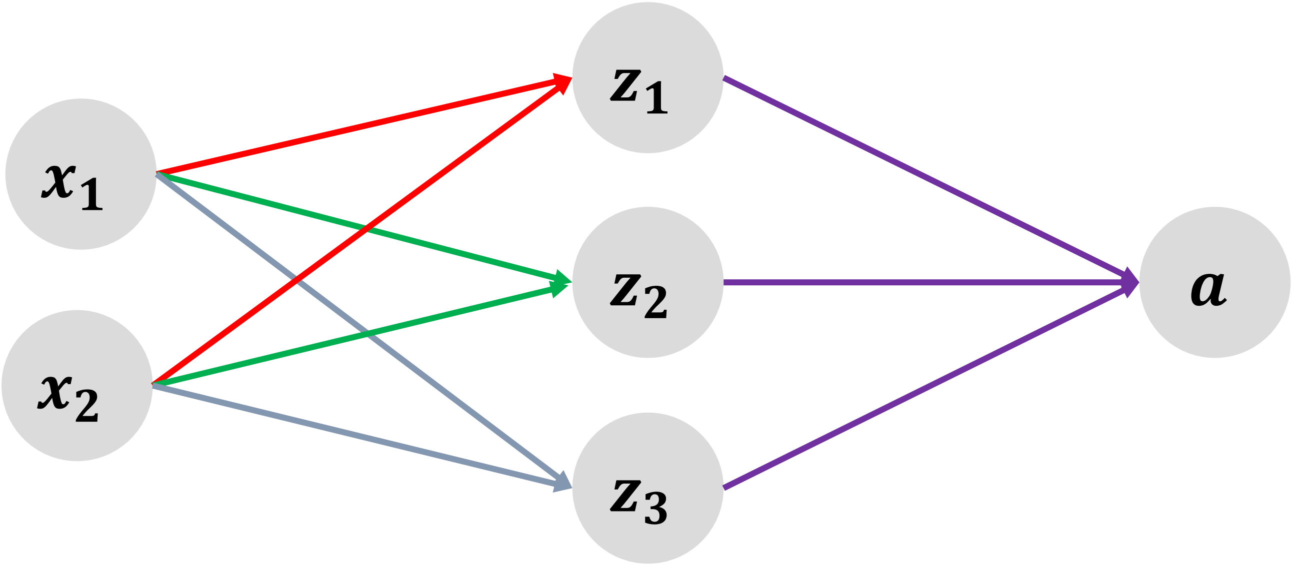
- Запишем наши вычисления в виде графа:



Вершины: вычисляемые переменные $(\boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{z_1}, \boldsymbol{z_2}, \boldsymbol{z_3}, \boldsymbol{a})$
Ребра: зависимости (нам нужен $\boldsymbol{x_1}$ и $\boldsymbol{x_2}$ чтобы получить $\boldsymbol{z_1}$)

У нашего графа вычислений есть имя

- Многослойный персептрон (MLP):



Входной слой **Скрытый слой** **Выходной слой**

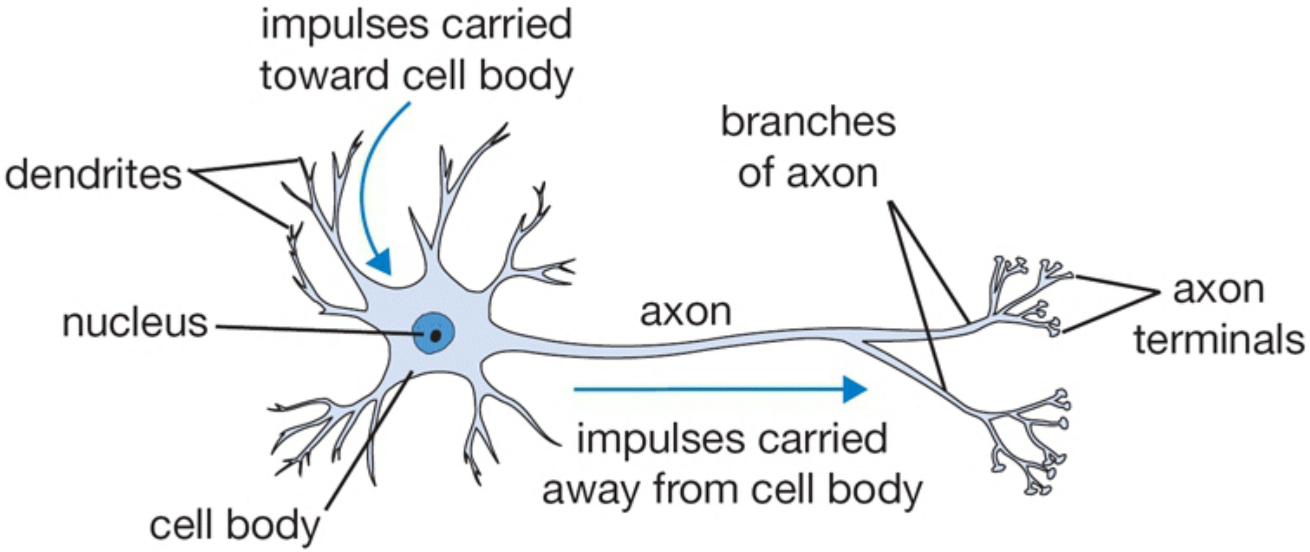
Признаки

Каждая вершина называется **нейроном**:

1. Линейная комбинация входов
2. Нелинейная функция **активации** (пример: $\sigma(x)$)

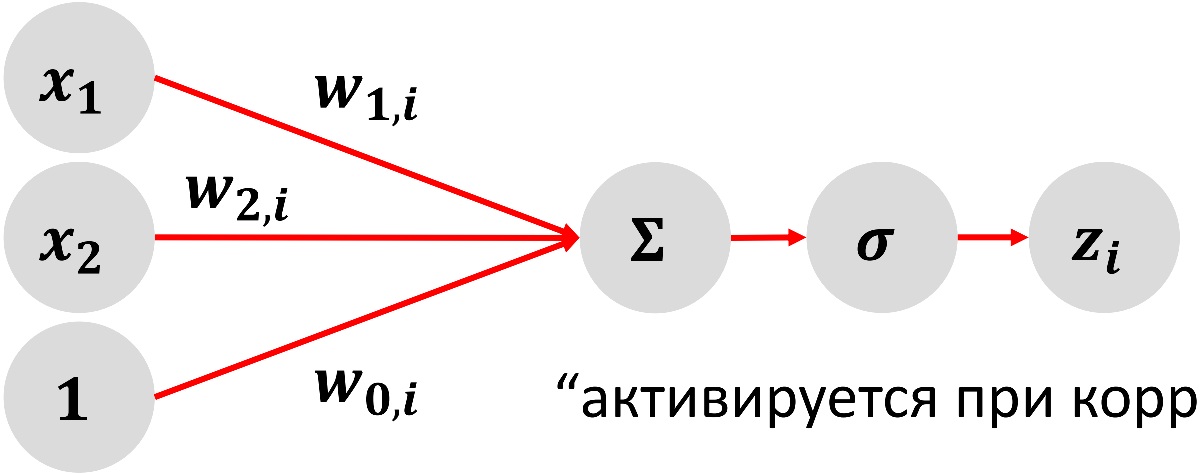
Почему нейрон?

- Нейрон человека:



$$z_i = \sigma(\boldsymbol{w_{0,i}} + \boldsymbol{w_{1,i}} x_1 + \boldsymbol{w_{2,i}} x_2)$$

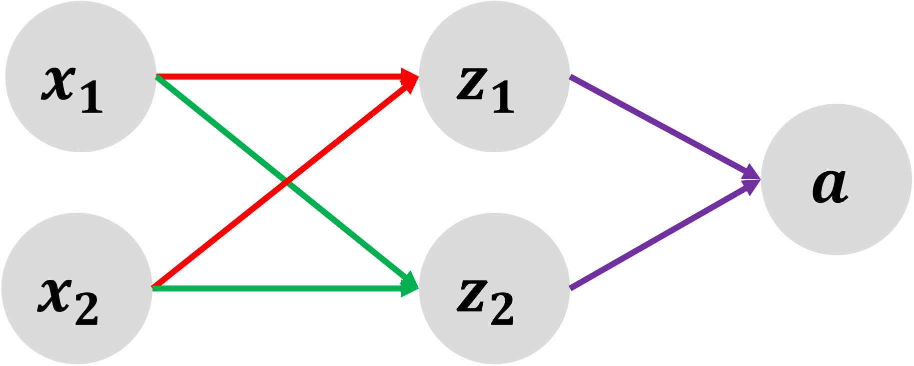
- Математический нейрон:



“активируется при корреляции”

Нам нужны нелинейности в нейронах!

- Давайте попробуем выкинуть $\sigma(x)$:

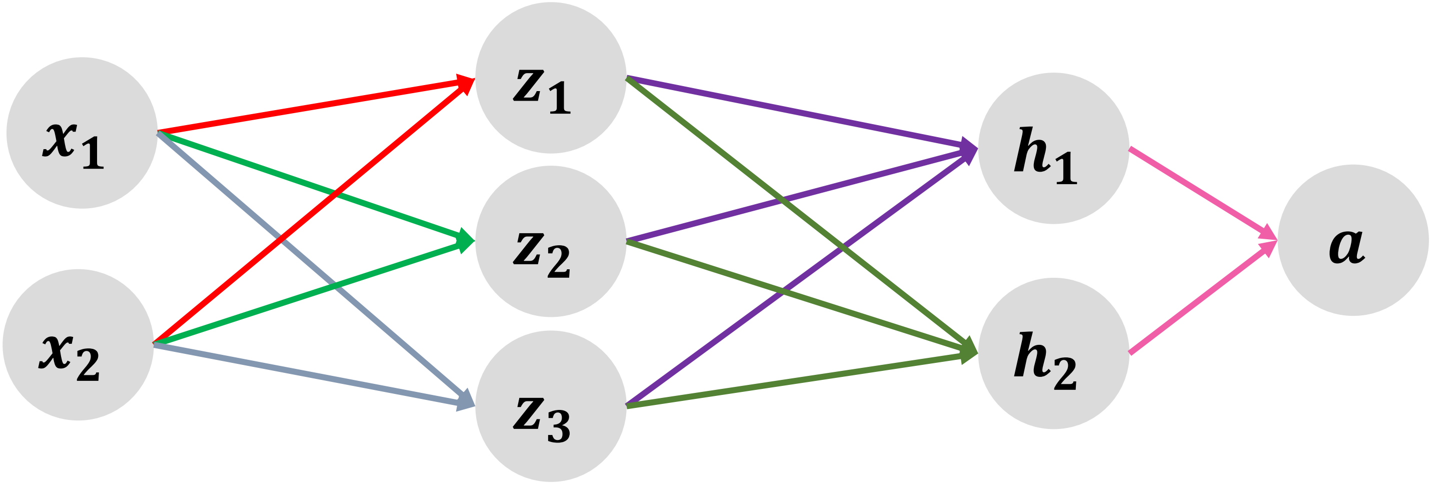


$$z_1 = w_{1,1}x_1 + w_{2,1}x_2$$
$$z_2 = w_{1,2}x_1 + w_{2,2}x_2$$
$$a = w_1 z_1 + w_2 z_2$$

- Наш алгоритм становится линейной функцией!
  - $a = (w_1 w_{1,1} + w_2 w_{1,2})x_1 + (w_1 w_{2,1} + w_2 w_{2,2})x_2$

Обзор MLP

- MLP – это простейший пример нейросети
- MLP может иметь много скрытых слоев:



- Архитектура MLP:
  - Кол-во слоев
  - Кол-во нейронов в каждом слое
  - Какую активацию использовать

- Скрытый слой в MLP называют:
  - Dense layer (плотный)
  - Fully-connected layer (полно-связный)

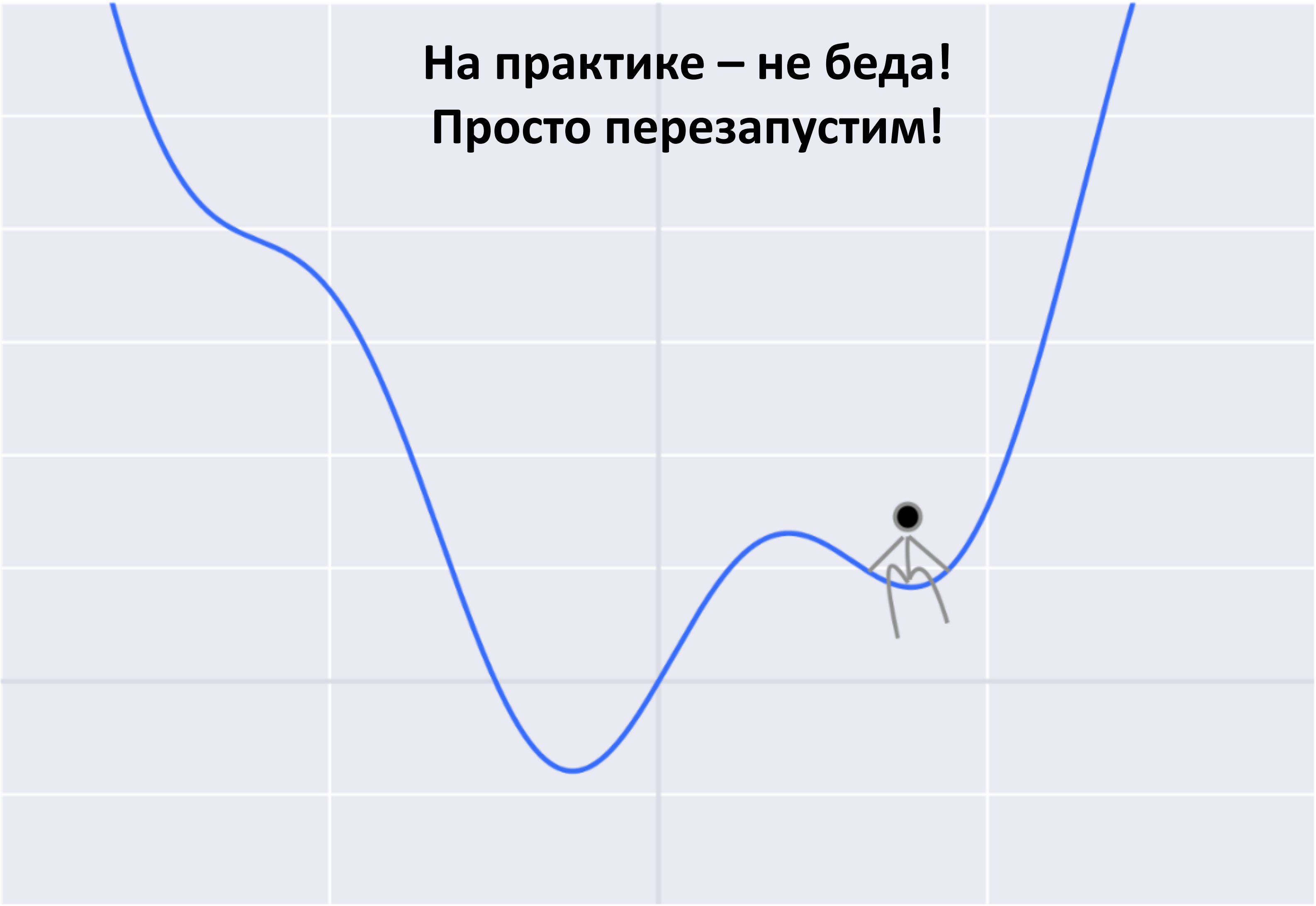
Хорошо, но как найти параметры MLP?

- Мы знаем как выучить параметры логистической регрессии – **градиентный спуск**

- Давайте здесь **сделаем то же самое**, ведь финальная функция дифференцируемая!

- Быстрый и эффективный способ вычисления градиента для любого дифференцируемого графа вычислений называется **back-propagation (обратное распространение ошибки)**

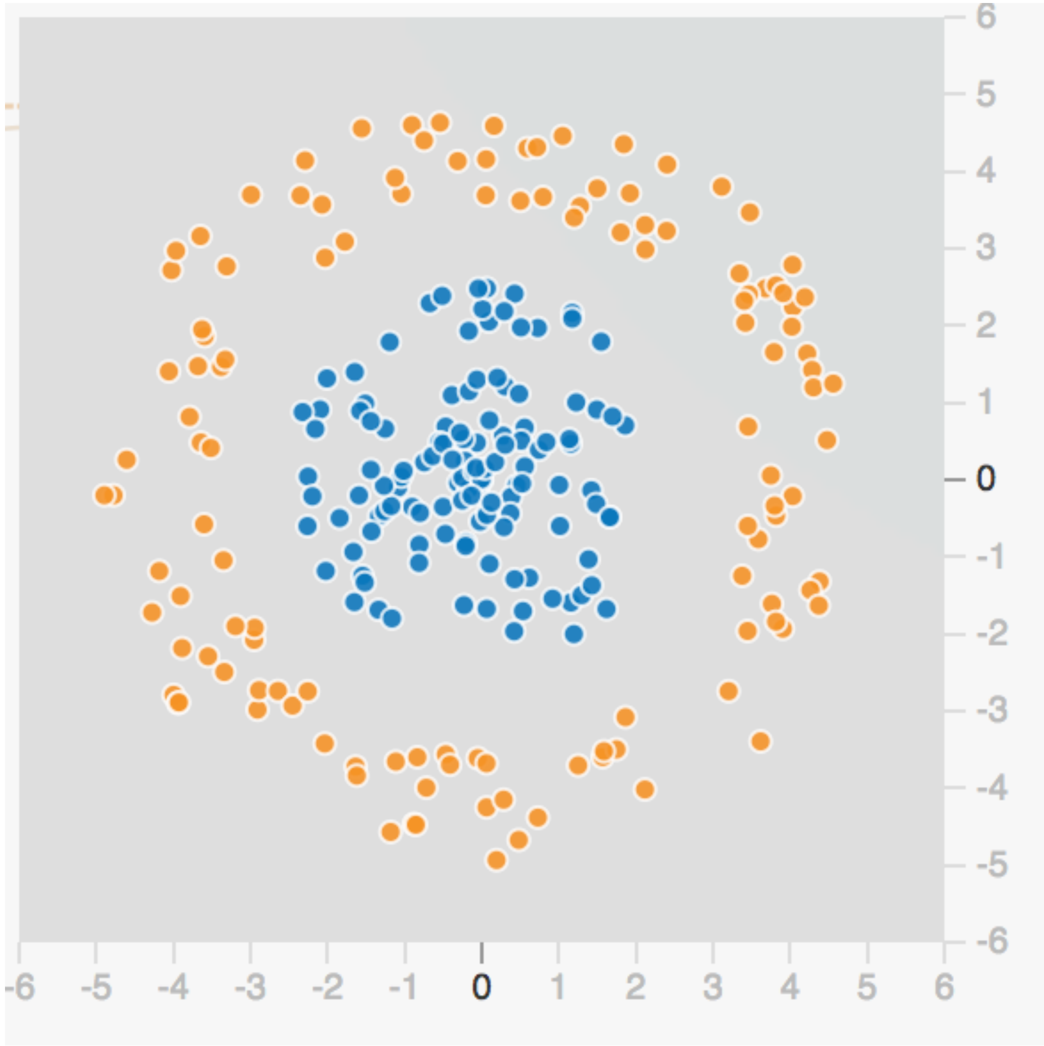
Застревает в локальных минимумах



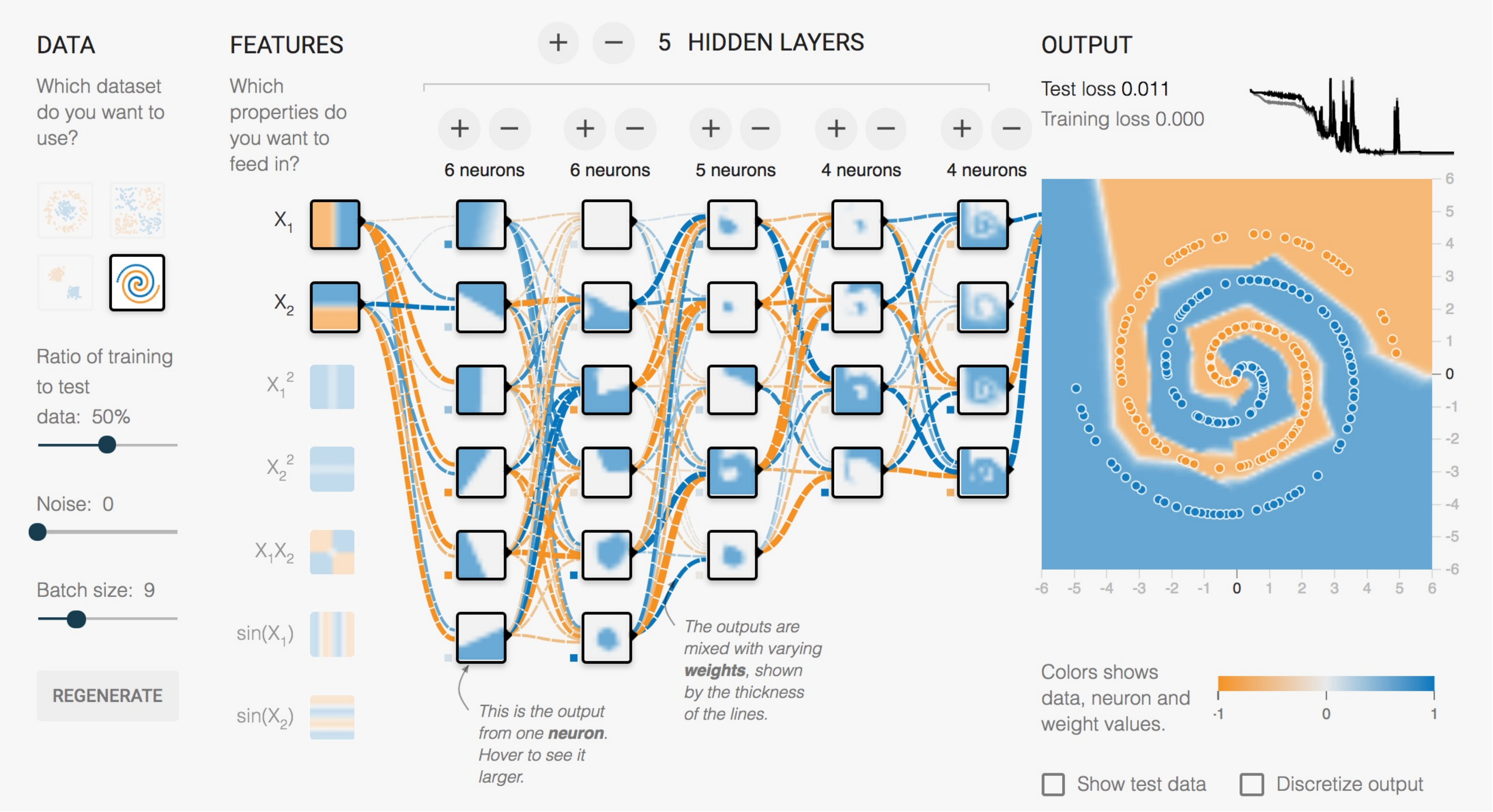
Bob chillin at a local optima

Демо: нейросети в TensorFlow Playground

- http://playground.tensorflow.org



Можно решать сложные задачи



* с функцией активации ReLU

Детали градиентного спуска: цепное правило

- Умеем дифференцировать: $\frac{dx^2}{dx} = 2x \quad \frac{de^x}{dx} = e^x \quad \frac{d\ln(x)}{dx} = \frac{1}{x}$

- Возьмем сложную функцию:

$$z_1 = z_1(x_1, x_2)$$

$$z_2 = z_2(x_1, x_2) \qquad \text{где } z_1, z_2, p \text{ дифференцируемы}$$

$$p = p(z_1, z_2)$$

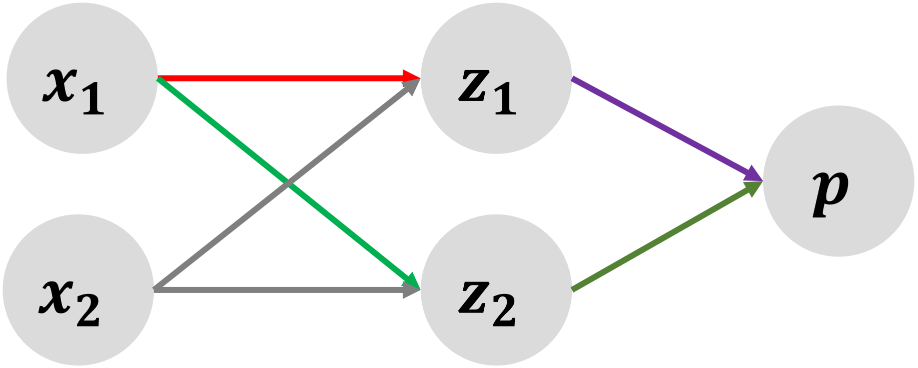
- Цепное правило:

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

- Пример для $h(x) = f(x)g(x)$: $\frac{\partial h}{\partial x} = \frac{\partial h}{\partial f}\frac{\partial f}{\partial x} + \frac{\partial h}{\partial g}\frac{\partial g}{\partial x} = g\frac{\partial f}{\partial x} + f\frac{\partial g}{\partial x}$

Граф для вычисления производных

- Граф для вычисления нашей композиции:

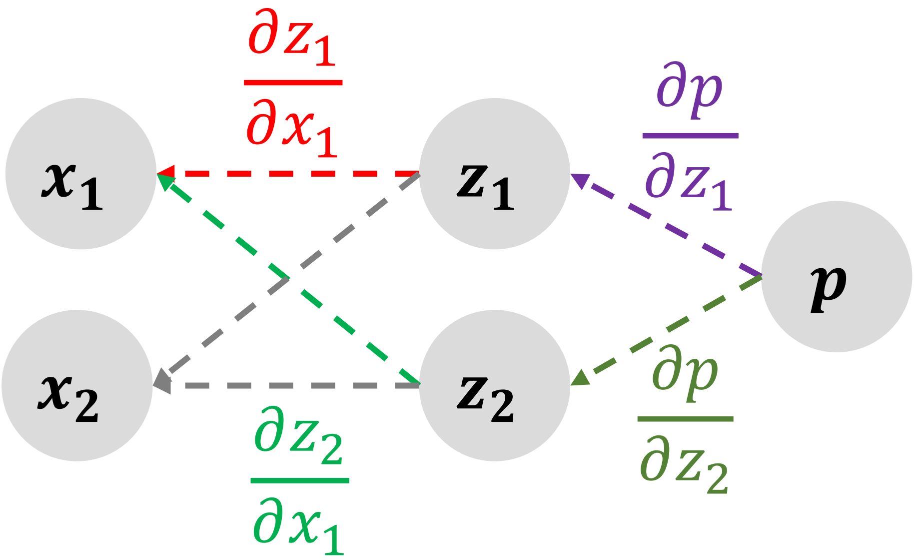


$$z_1 = z_1(x_1, x_2)$$

$$z_2 = z_2(x_1, x_2)$$

$$p = p(z_1, z_2)$$

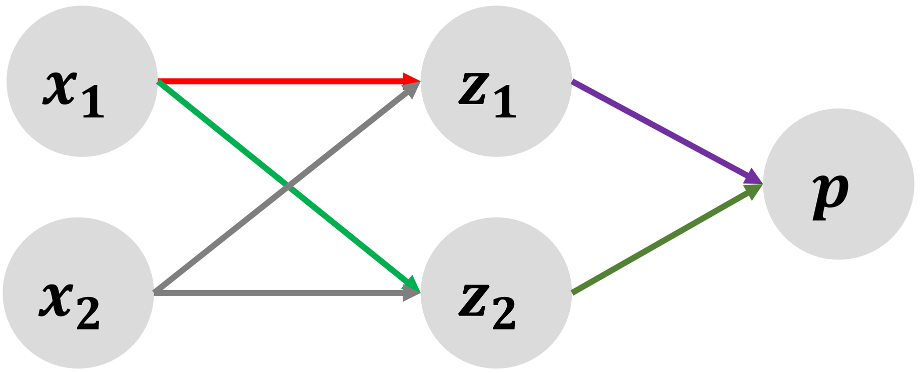
- Построим из него граф из производных:



Каждому ребру приписана производная начала по концу

Граф для вычисления производных

- Граф для вычисления нашей композиции:

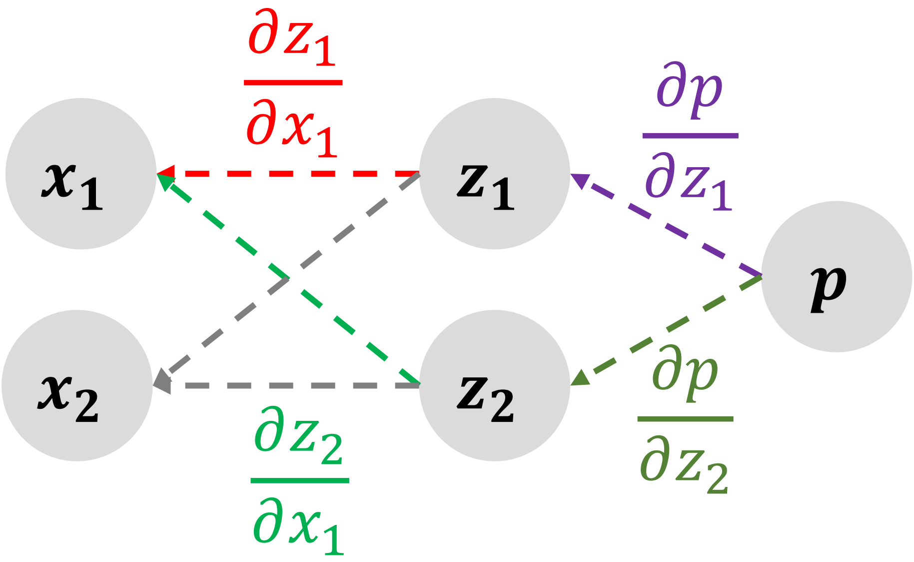


$$z_1 = z_1(x_1, x_2)$$

$$z_2 = z_2(x_1, x_2)$$

$$p = p(z_1, z_2)$$

- Построим из него граф из производных:

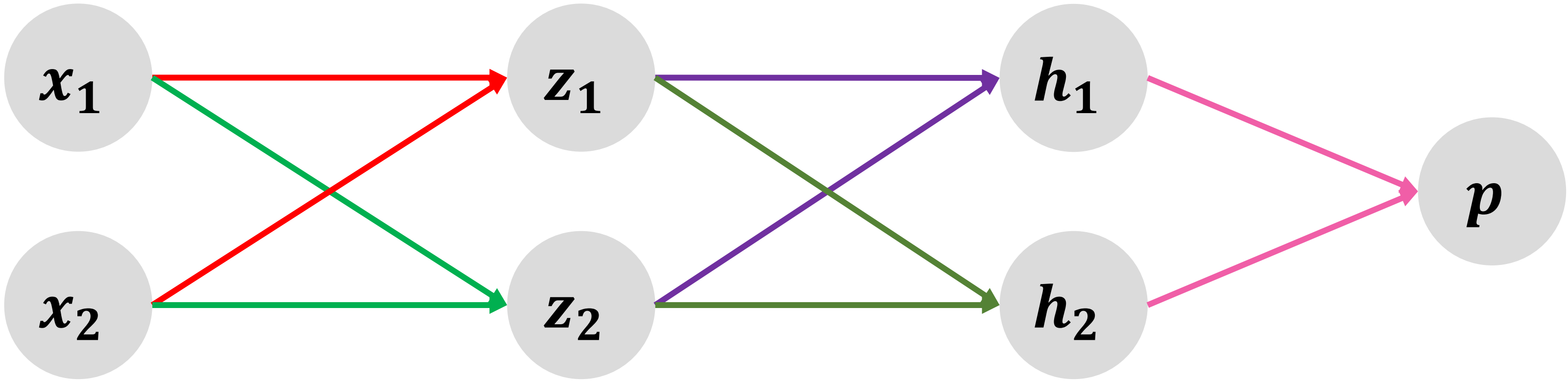


$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

Можно догадаться как работает **цепное правило**

Пойдем глубже

- Добавим еще один скрытый слой:



$$z_1 = z_1(x_1, x_2) \qquad h_1 = h_1(z_1, z_2)$$

$$z_2 = z_2(x_1, x_2) \qquad h_2 = h_2(z_1, z_2)$$

$$p = p(h_1, h_2)$$

Пойдем глубже

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial x_1}$$

Несколько раз применим цепное правило!

Пойдем глубже

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\boxed{\frac{\partial h_1}{\partial x_1}} + \frac{\partial p}{\partial h_2}\boxed{\frac{\partial h_2}{\partial x_1}}$$

$$\boxed{\frac{\partial h_1}{\partial x_1} = \frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

$$\boxed{\frac{\partial h_2}{\partial x_1} = \frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

$$z_1 = z_1(x_1, x_2) \qquad h_1 = h_1(z_1, z_2)$$

$$z_2 = z_2(x_1, x_2) \qquad h_2 = h_2(z_1, z_2)$$

Пойдем глубже

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\boxed{\frac{\partial h_1}{\partial x_1}} + \frac{\partial p}{\partial h_2}\boxed{\frac{\partial h_2}{\partial x_1}}$$

$$\boxed{\frac{\partial h_1}{\partial x_1} = \frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

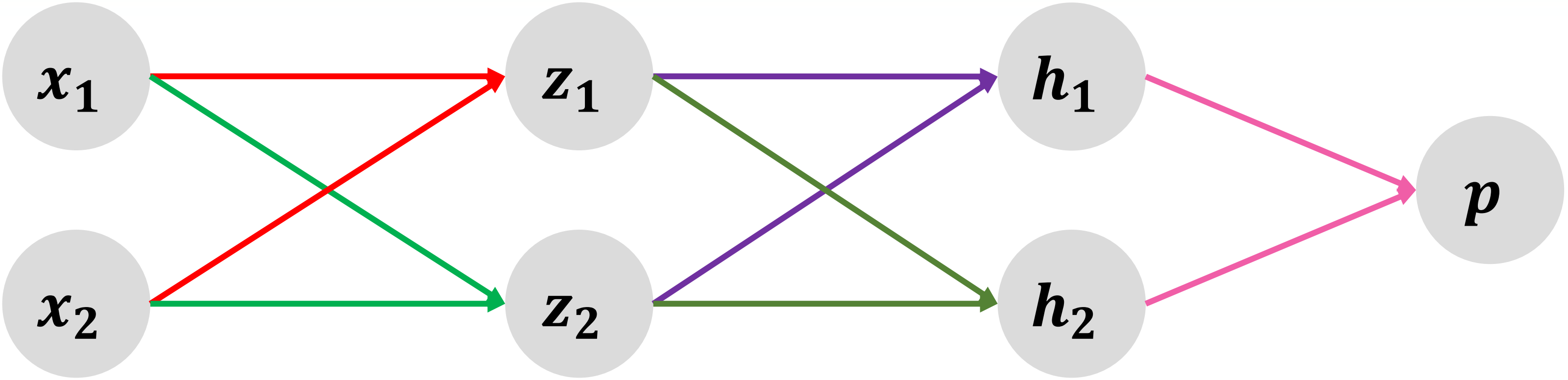
$$\boxed{\frac{\partial h_2}{\partial x_1} = \frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\boxed{\left(\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}\right)} + \frac{\partial p}{\partial h_2}\boxed{\left(\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}\right)}$$

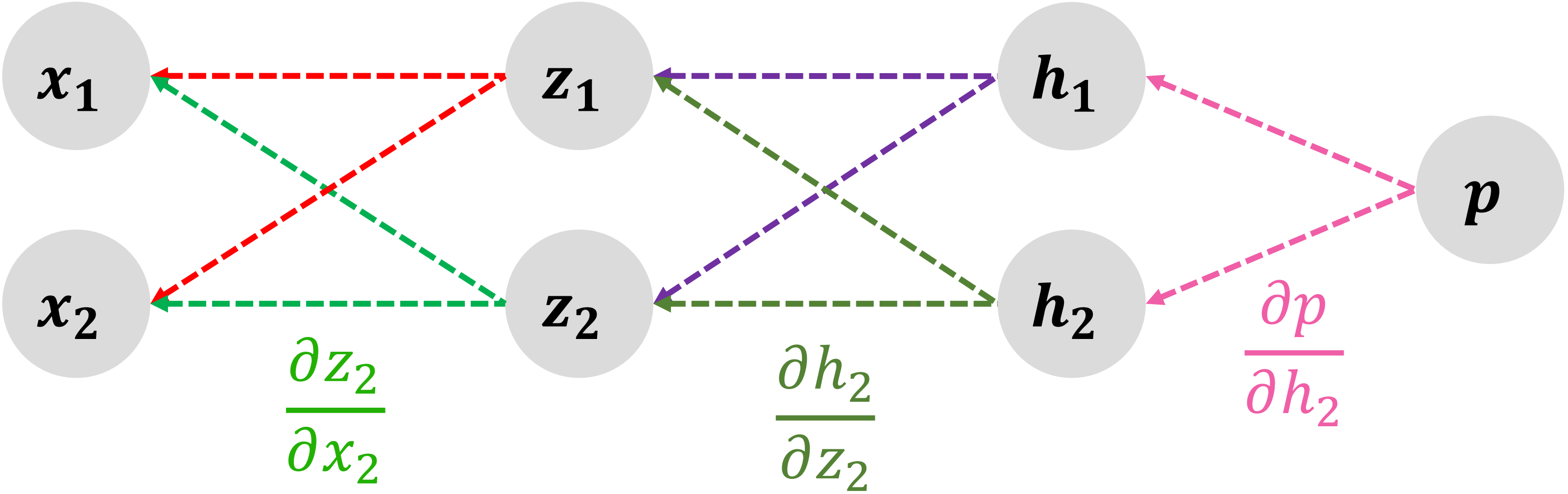
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

Проверим граф производных!

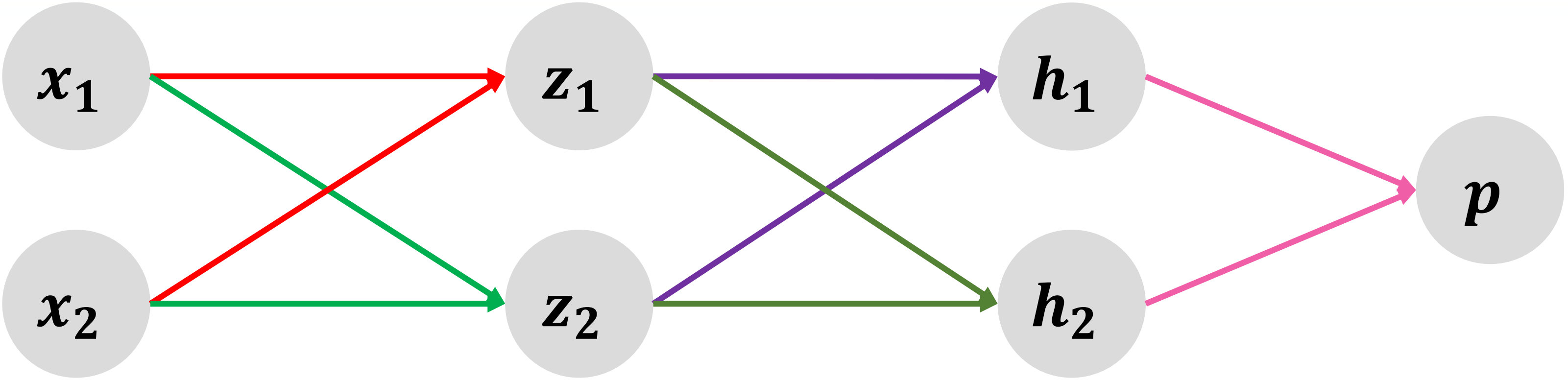
Пойдем глубже



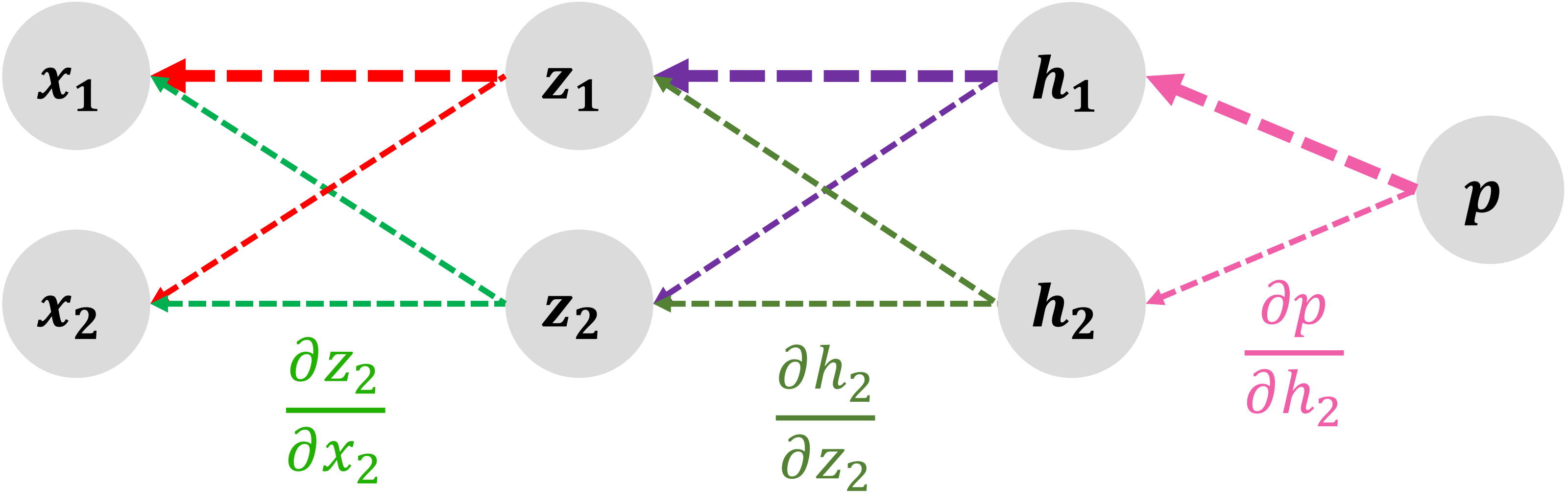
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$



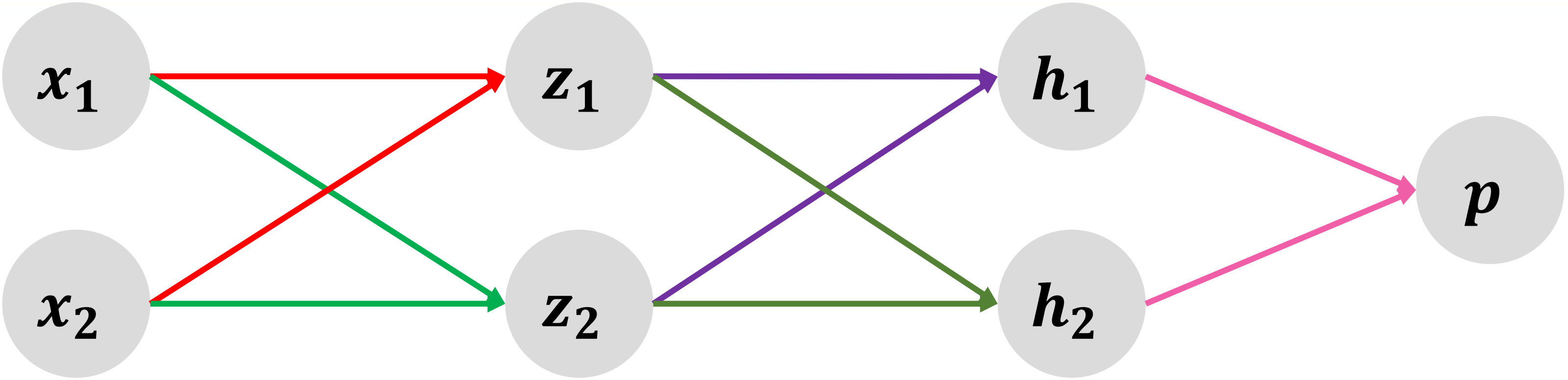
Пойдем глубже



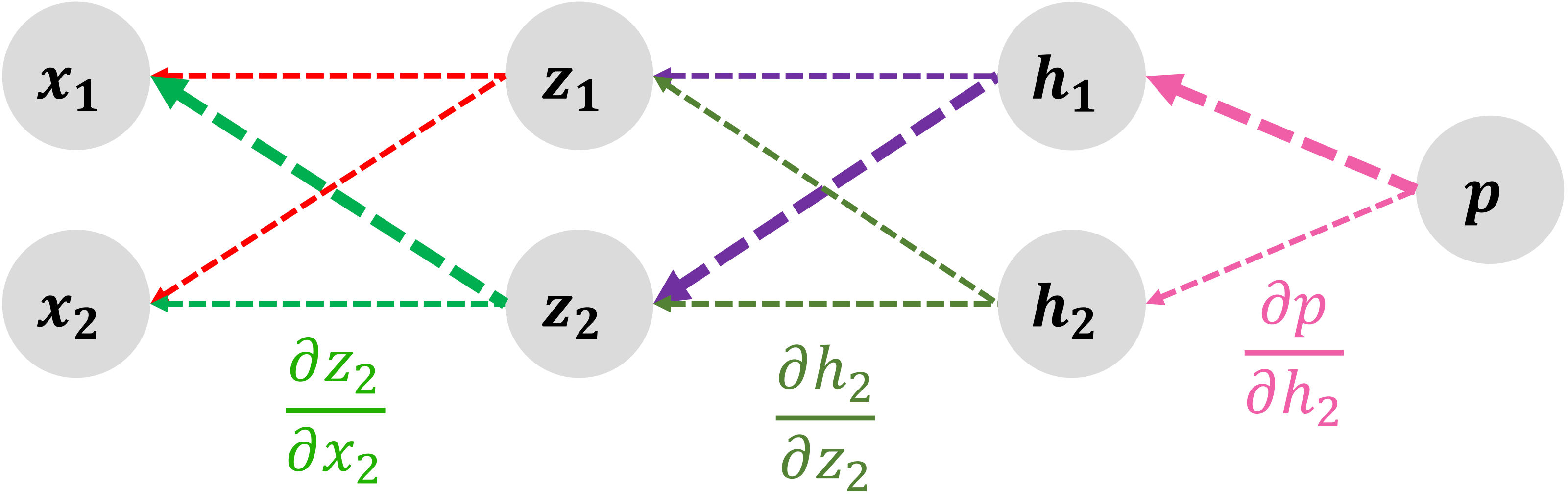
$$\frac{\partial p}{\partial x_1} = \boxed{\frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1}} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$



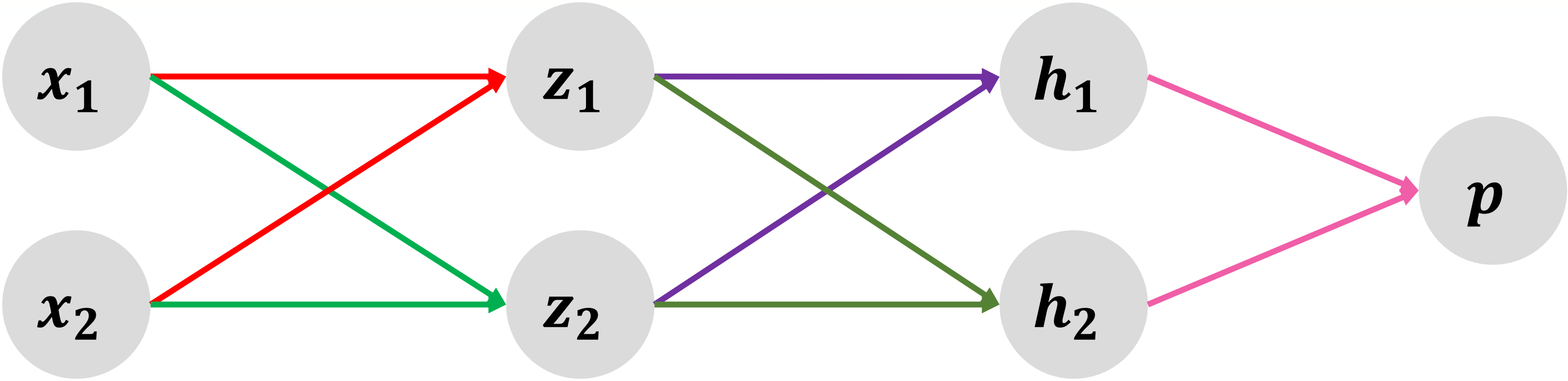
Пойдем глубже



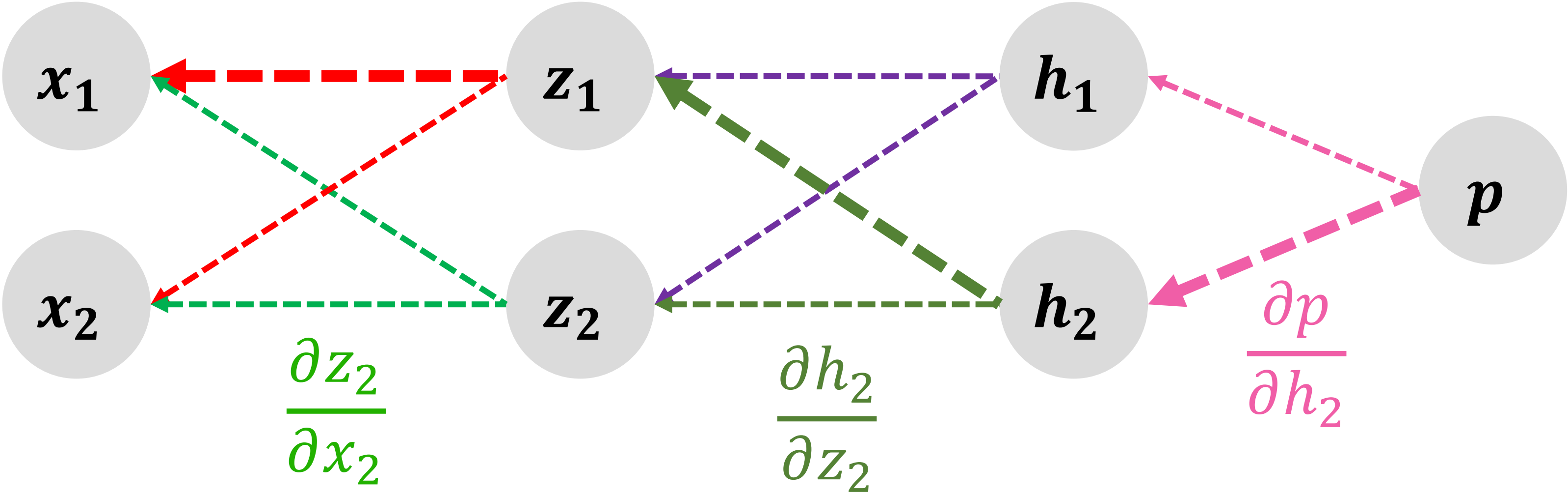
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \boxed{\frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1}} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$



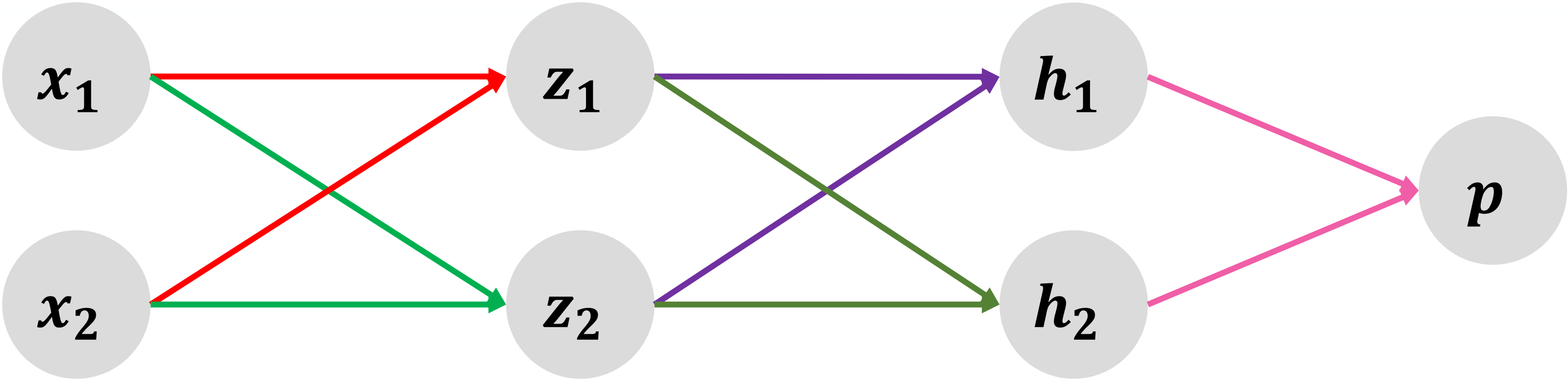
Пойдем глубже



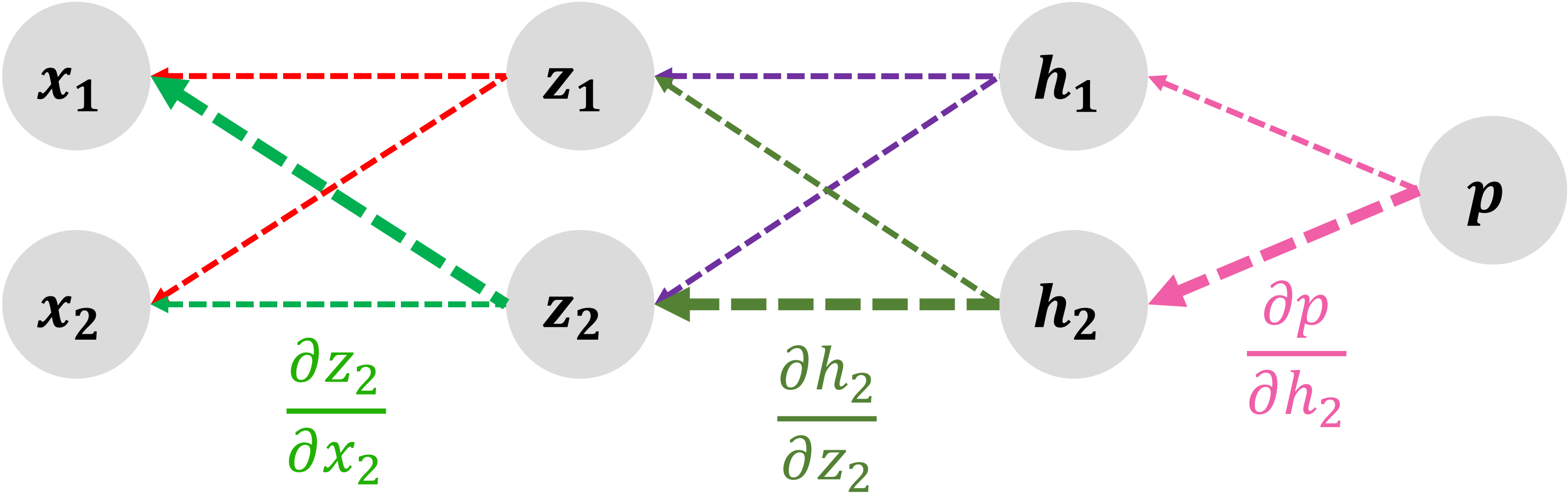
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \boxed{\frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1}} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$



Пойдем глубже



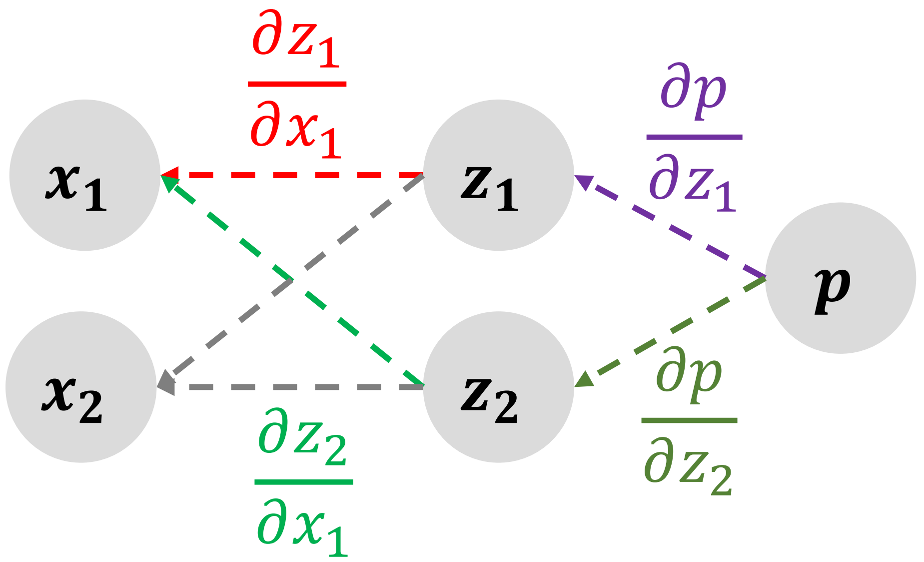
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \boxed{\frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}}$$



Алгоритм вычисления производной в графе

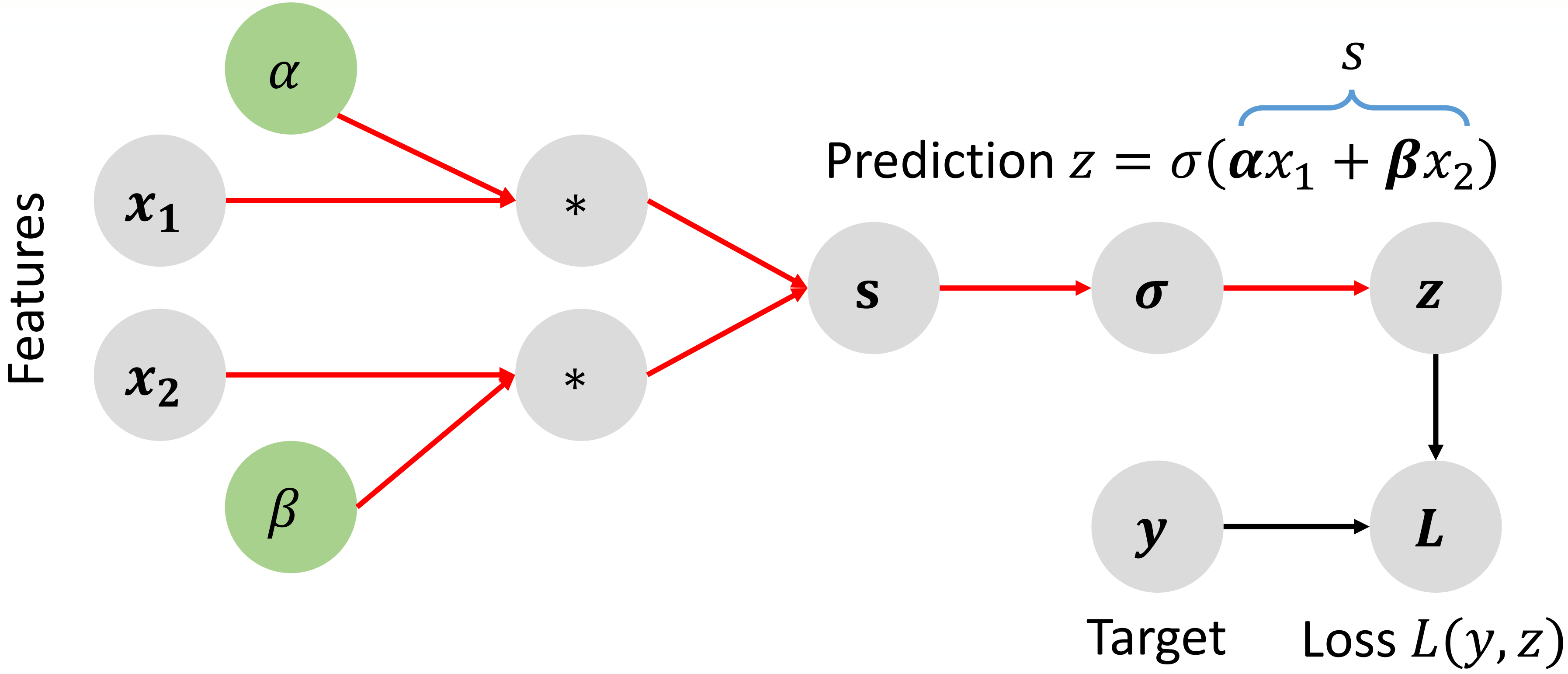
Как посчитать производную a по b:

- Находим непосещенный путь из a в b
- Перемножаем значения на ребрах в пути
- Добавляем в сумму



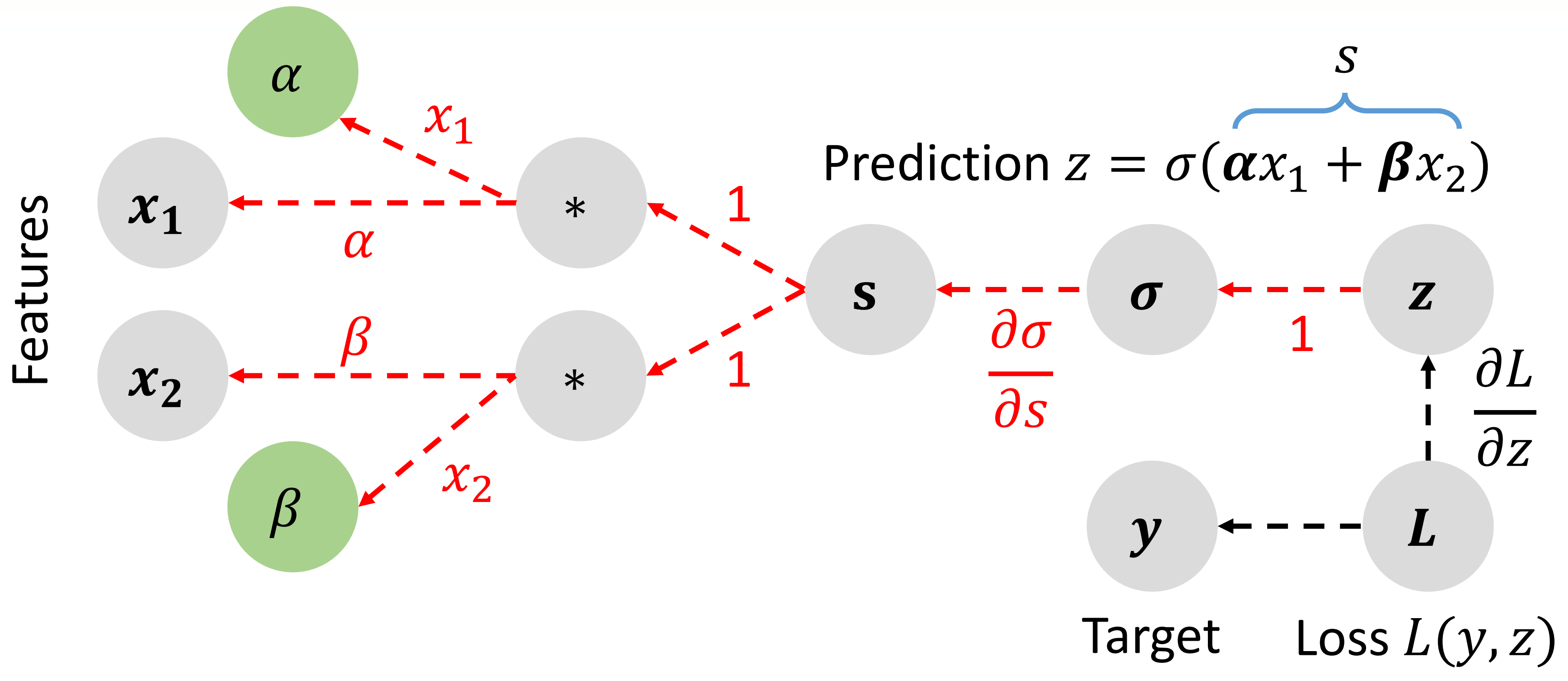
$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

На примере одного нейрона



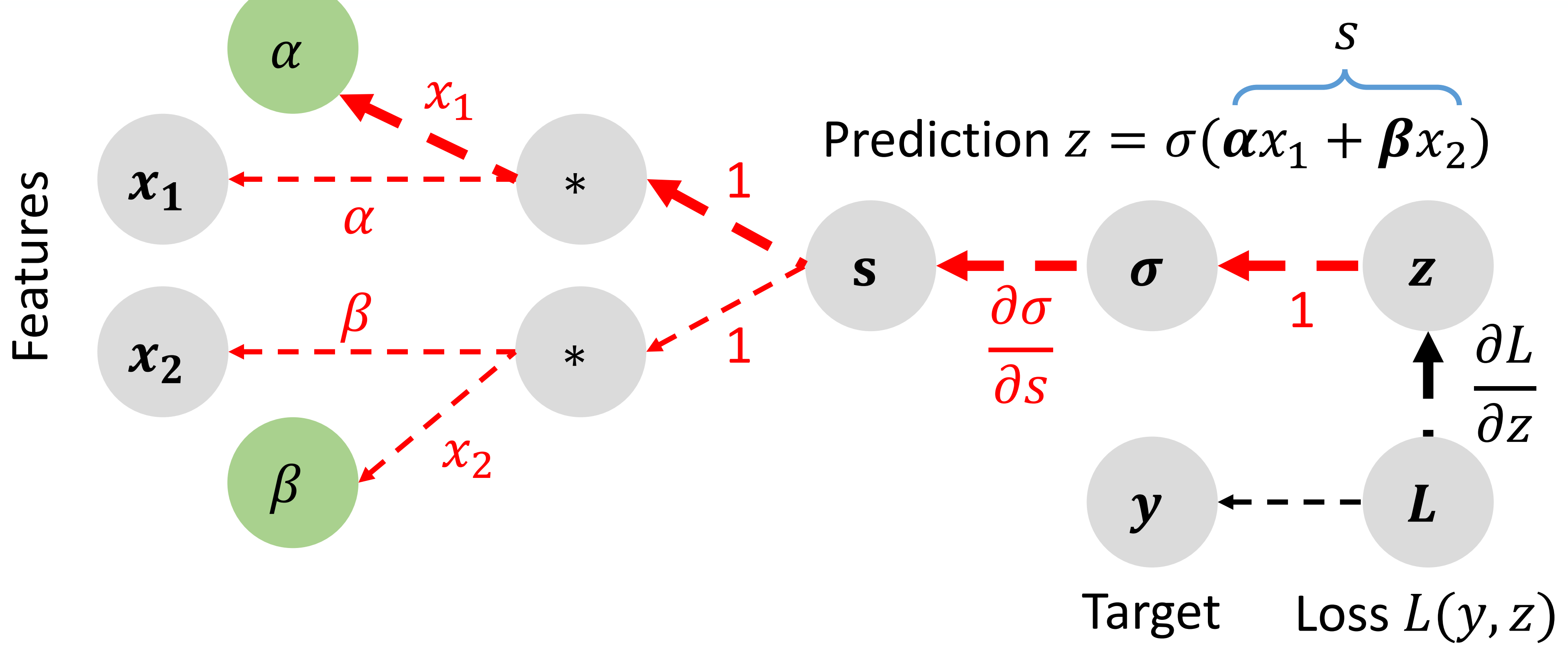
Для SGD нам нужны $\dfrac{\partial L}{\partial \alpha}, \dfrac{\partial L}{\partial \beta}$

Граф вычисления производных



Для SGD нам нужны $\dfrac{\partial L}{\partial \alpha}, \dfrac{\partial L}{\partial \beta}$

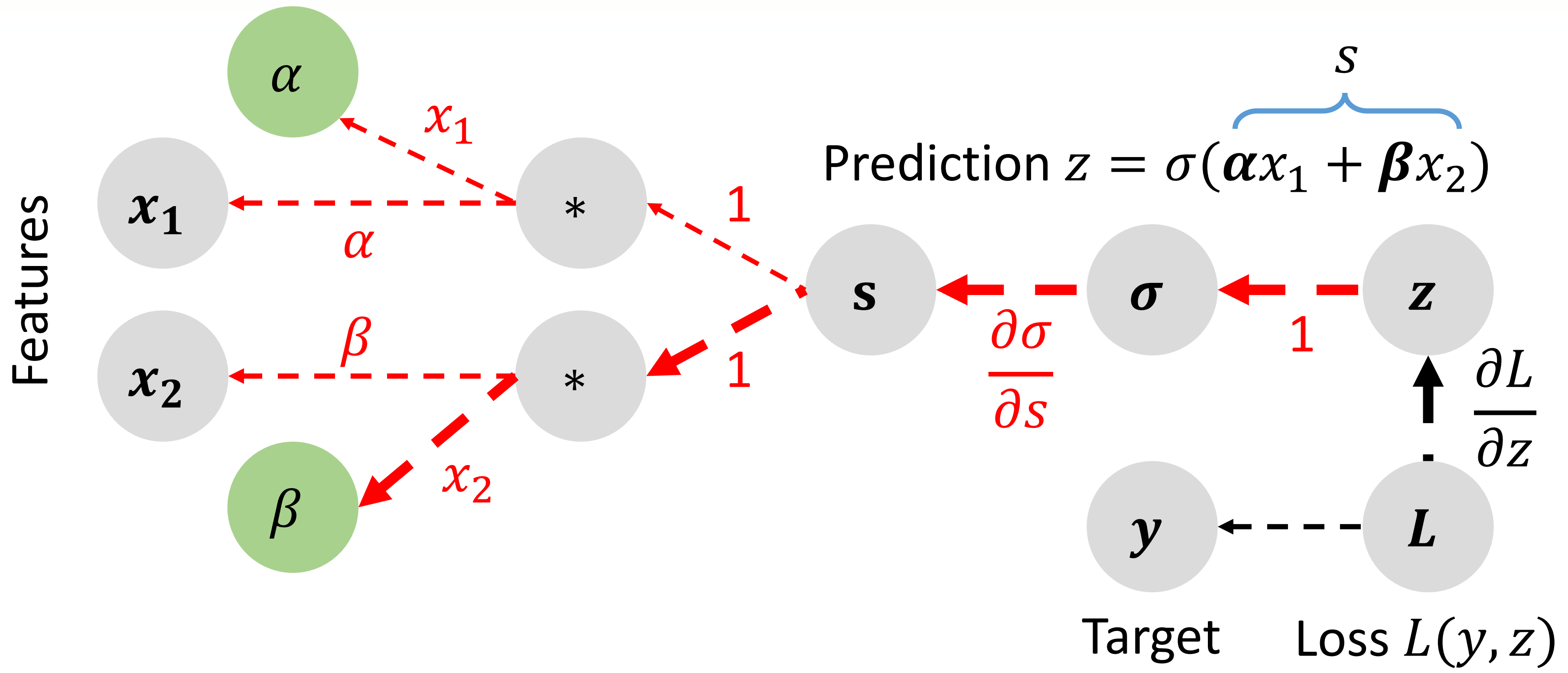
Граф вычисления производных



Для SGD нам нужны $\frac{\partial L}{\partial \alpha}, \frac{\partial L}{\partial \beta}$

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial z} \frac{\partial \sigma}{\partial s} x_1$$

Граф вычисления производных



Для SGD нам нужны $\dfrac{\partial L}{\partial \alpha}, \dfrac{\partial L}{\partial \beta}$

$$\frac{\partial L}{\partial \alpha} = \frac{\partial L}{\partial z}\frac{\partial \sigma}{\partial s} x_1 \qquad \frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial z}\frac{\partial \sigma}{\partial s} x_2$$

Цепное правило и граф производных

- Теперь у нас есть алгоритм для подсчета производных для любых дифференцируемых графов вычислений
- Осталось делать вычисления быстро!

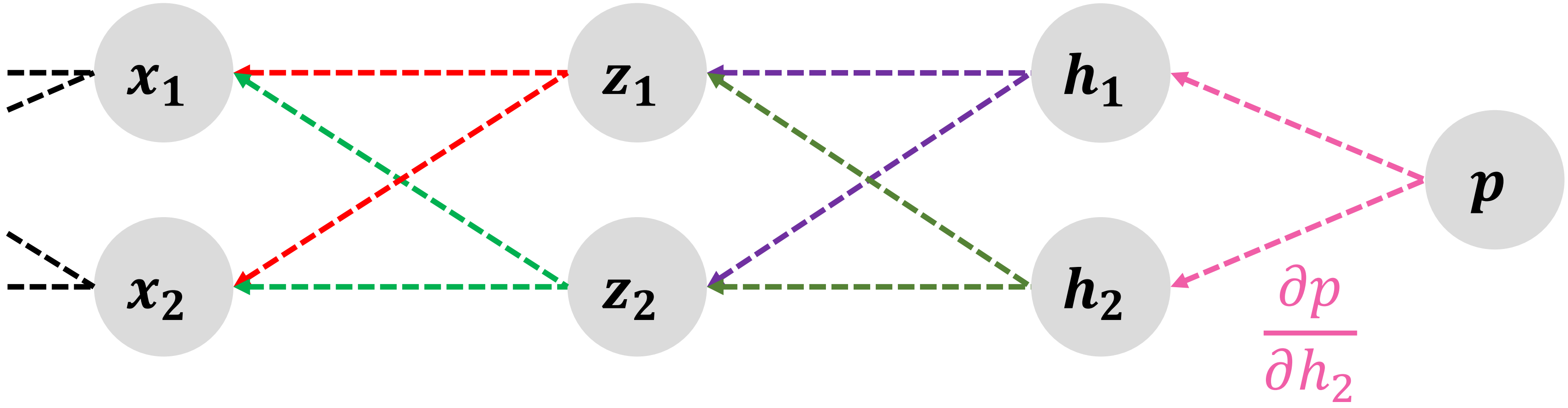
MLP с 3 скрытыми слоями

На самом деле мы хотим менять параметры нейрона:

$$h_2 = \sigma(\boldsymbol{w_0} + \ \boldsymbol{w_1} z_1 + \boldsymbol{w_2} z_2)$$

Градиент: $$\frac{\partial L}{\partial w_1} = \frac{\partial L}{\partial p} \frac{\partial p}{\partial w_1} = \frac{\partial L}{\partial p} \frac{\partial p}{\partial h_2} \frac{\partial h_2}{\partial w_1}$$

Для упрощения выкладок не будем копать вглубь нейронов!



Вот это нужно для SGD

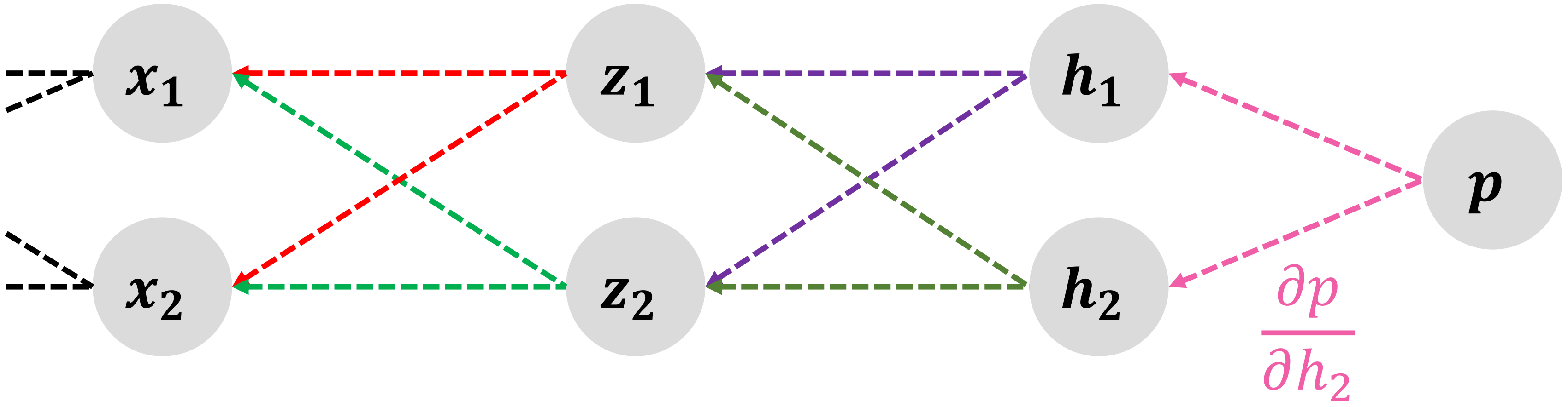
3: $\frac{\partial p}{\partial h_1}$ $\frac{\partial p}{\partial h_2}$

2: $\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}$ $\frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$

1:

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$\frac{\partial p}{\partial x_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



Вот это нужно для SGD

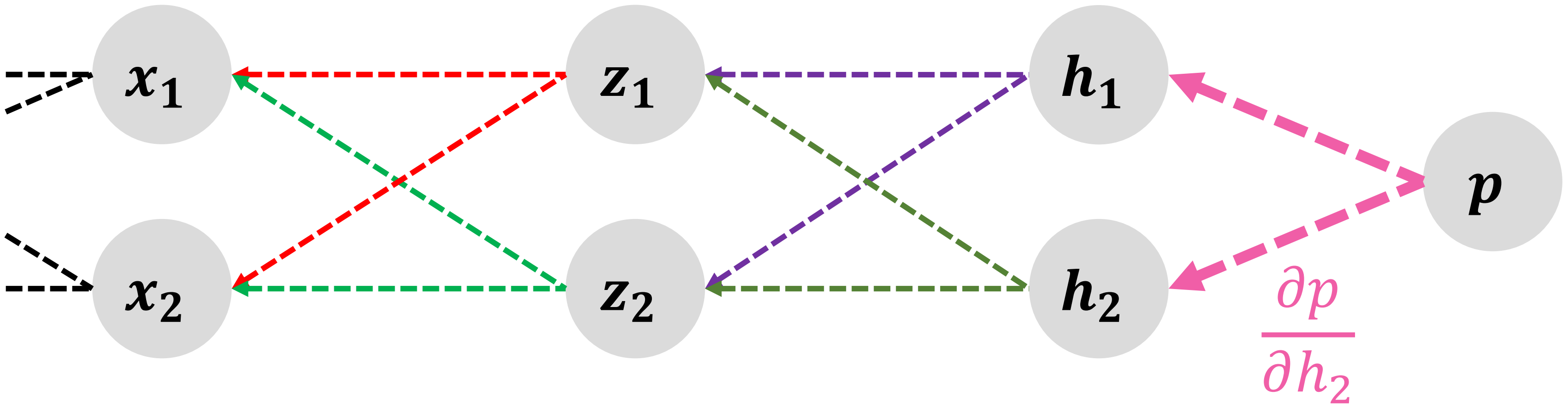
3: $\boxed{\frac{\partial p}{\partial h_1}}$ $\boxed{\frac{\partial p}{\partial h_2}}$

2: $\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}$ $\frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$

1:

$$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$\frac{\partial p}{\partial x_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



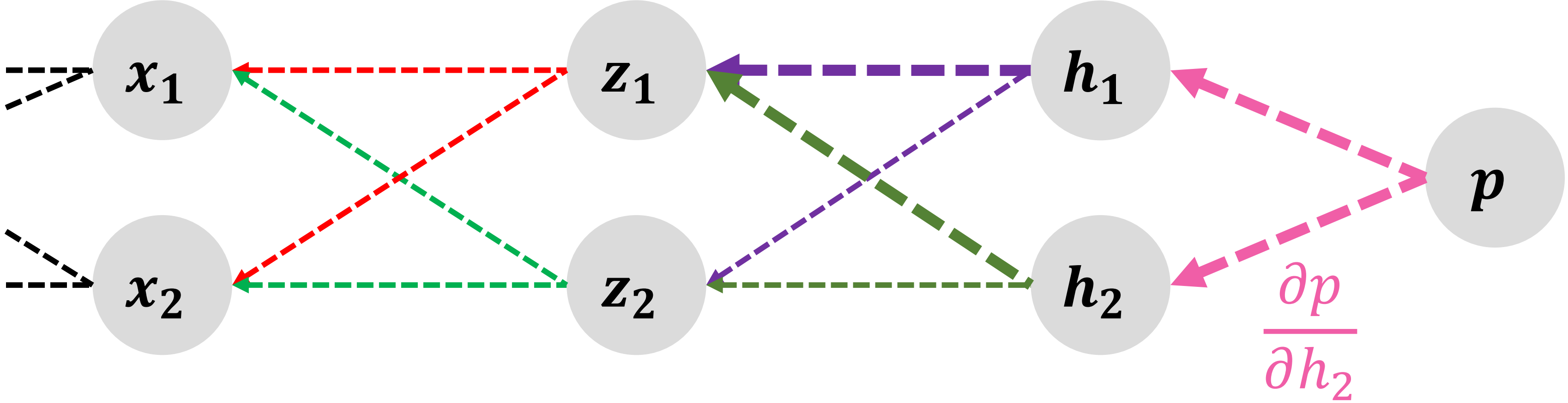
Вот это нужно для SGD

3: $\frac{\partial p}{\partial h_1}$ $\qquad$ $\frac{\partial p}{\partial h_2}$

2: $$\boxed{\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}} \qquad \frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$$

1: $$\frac{\partial p}{\partial x_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$\frac{\partial p}{\partial x_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



Вот это нужно для SGD

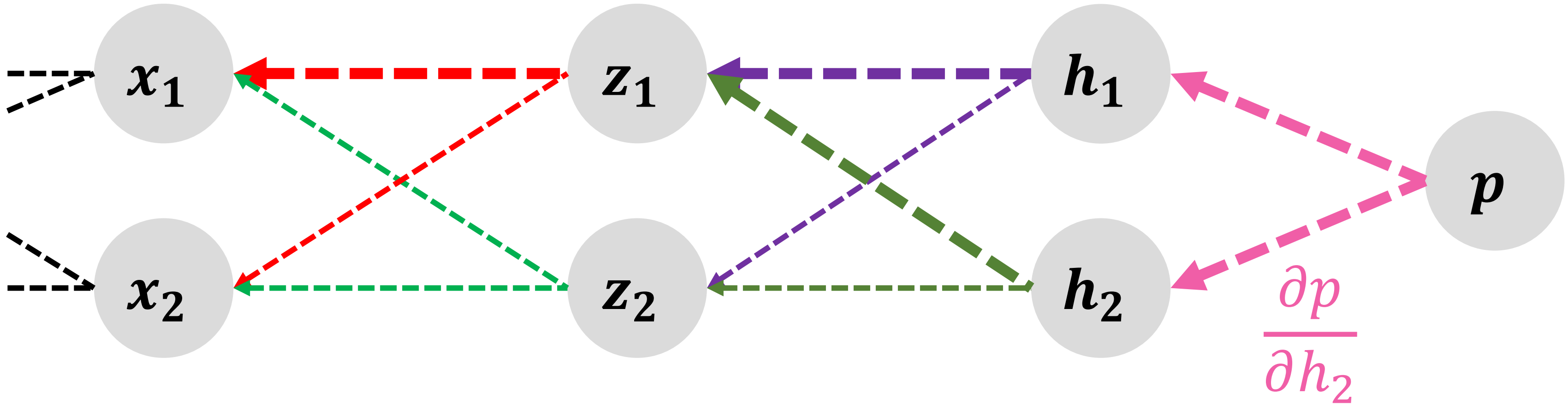
3: $\frac{\partial p}{\partial h_1}$ $\quad \frac{\partial p}{\partial h_2}$

2: $\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}$ $\quad \frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$

1:

$$\frac{\partial p}{\partial x_1} = \boxed{\frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_1}} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$\frac{\partial p}{\partial x_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



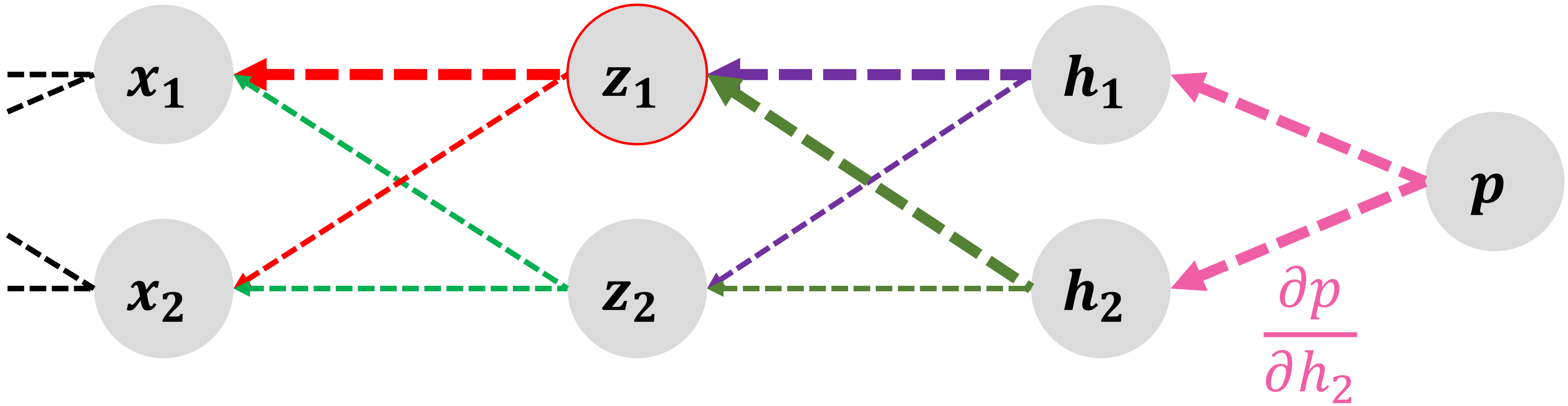
Вот это нужно для SGD

3: $\frac{\partial p}{\partial h_1}$ $\frac{\partial p}{\partial h_2}$

2: $\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}$ $\frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$

1:
$$\frac{\partial p}{\partial x_1} = \boxed{\left(\frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\right)\frac{\partial z_1}{\partial x_1}} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$\frac{\partial p}{\partial x_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



Вот это нужно для SGD

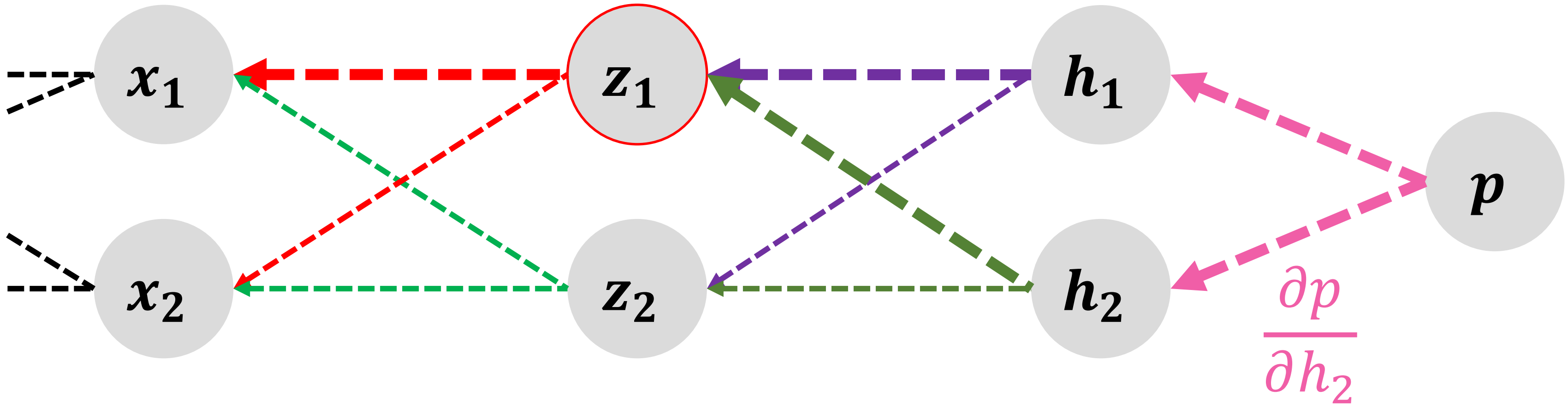
3: $\frac{\partial p}{\partial h_1}$ $\frac{\partial p}{\partial h_2}$

2: $\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}$ $\frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$

1:

$$\frac{\partial p}{\partial x_1} = \boxed{\left(\frac{\partial p}{\partial z_1}\right)\frac{\partial z_1}{\partial x_1}} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$

$$\frac{\partial p}{\partial x_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}\frac{\partial z_1}{\partial x_2} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$

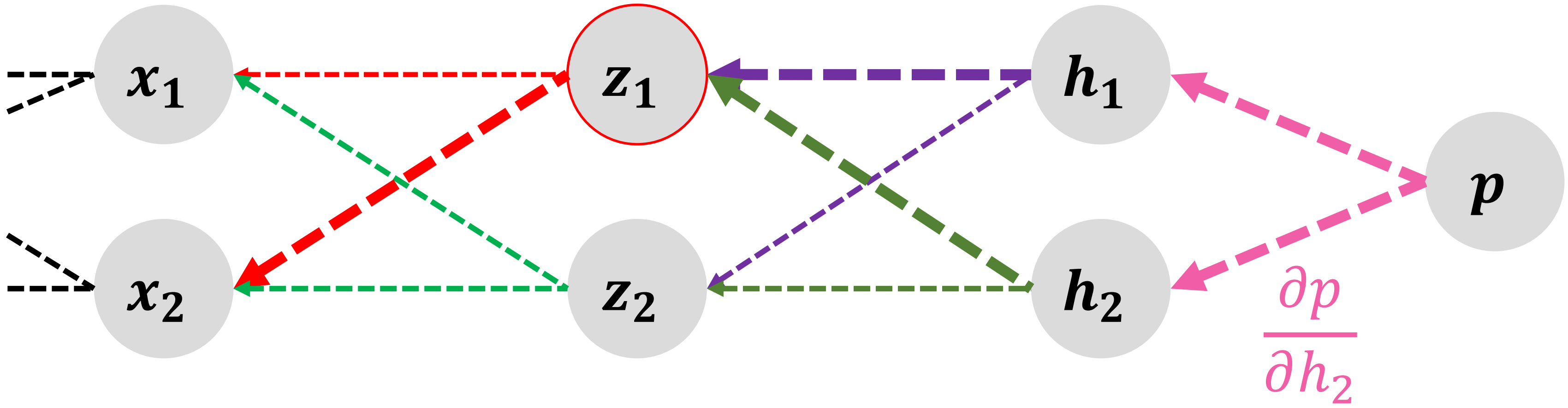


Вот это нужно для SGD

3: $\frac{\partial p}{\partial h_1}$ $\frac{\partial p}{\partial h_2}$

2: $\frac{\partial p}{\partial z_1} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_1}$ $\frac{\partial p}{\partial z_2} = \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}$

1:
$$\frac{\partial p}{\partial x_1} = \left(\frac{\partial p}{\partial z_1}\right)\frac{\partial z_1}{\partial x_1} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_1} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_1}$$
$$\frac{\partial p}{\partial x_2} = \boxed{\left(\frac{\partial p}{\partial z_1}\right)\frac{\partial z_1}{\partial x_2}} + \frac{\partial p}{\partial h_1}\frac{\partial h_1}{\partial z_2}\frac{\partial z_2}{\partial x_2} + \frac{\partial p}{\partial h_2}\frac{\partial h_2}{\partial z_2}\frac{\partial z_2}{\partial x_2}$$



Это называется reverse-mode дифференцирование

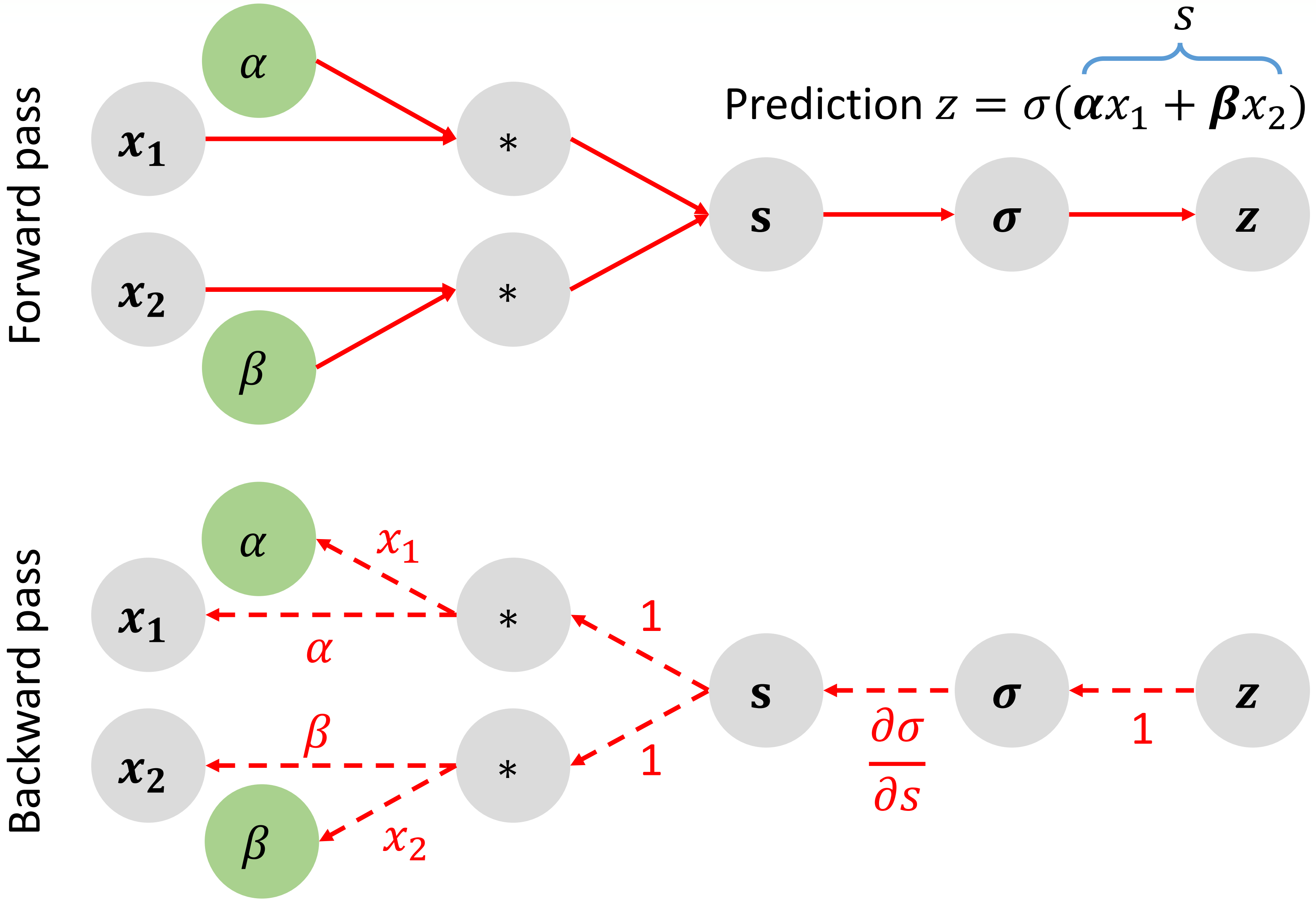
- В теории нейросетей это называют back-propagation (обратное распространение ошибки)

- Работает быстро, потому что переиспользует вычисленные ранее значения

- На самом деле, по каждому ребру пройдемся всего раз, то есть сложность линейна по количеству ребер (т.е. параметров)!

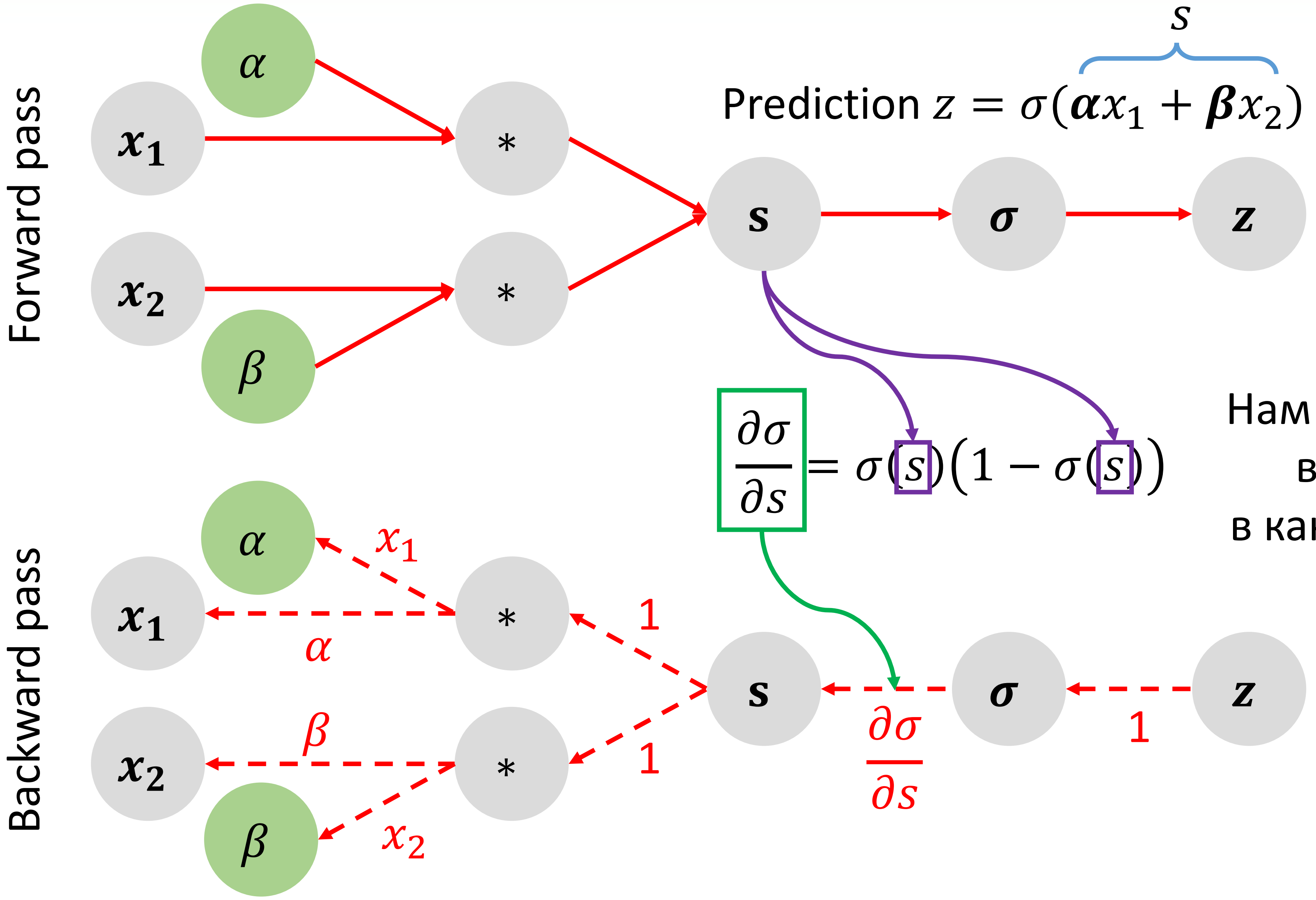
Back-propagation (Back-prop)

Prediction $z = \sigma(\overbrace{\boldsymbol{\alpha} x_1 + \boldsymbol{\beta} x_2}^{s})$



Back-propagation (Back-prop)

Prediction $z = \sigma(\overbrace{\boldsymbol{\alpha} x_1 + \boldsymbol{\beta} x_2}^{s})$

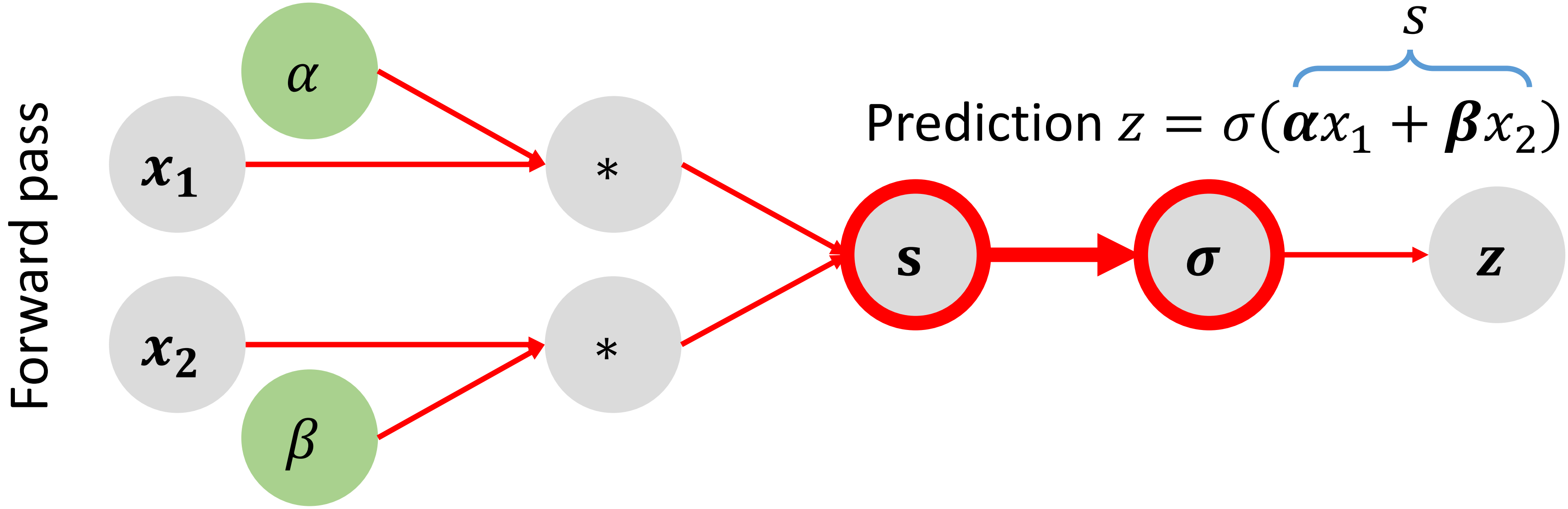


$$\boxed{\frac{\partial \sigma}{\partial s}} = \sigma(\boxed{s})(1 - \sigma(\boxed{s}))$$

Нам же нужно ее вычислить в какой-то точке!

Интерфейс прямого прохода

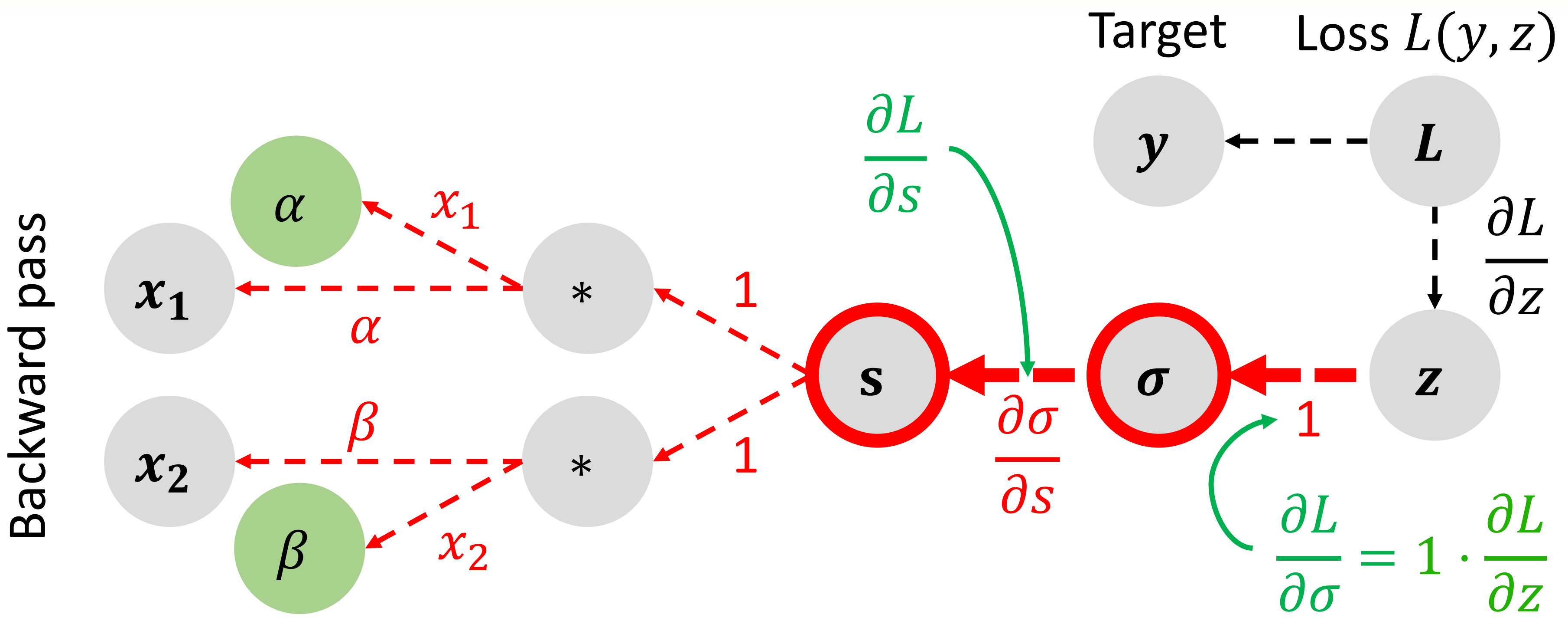
Реализуем сигмоидную активацию!



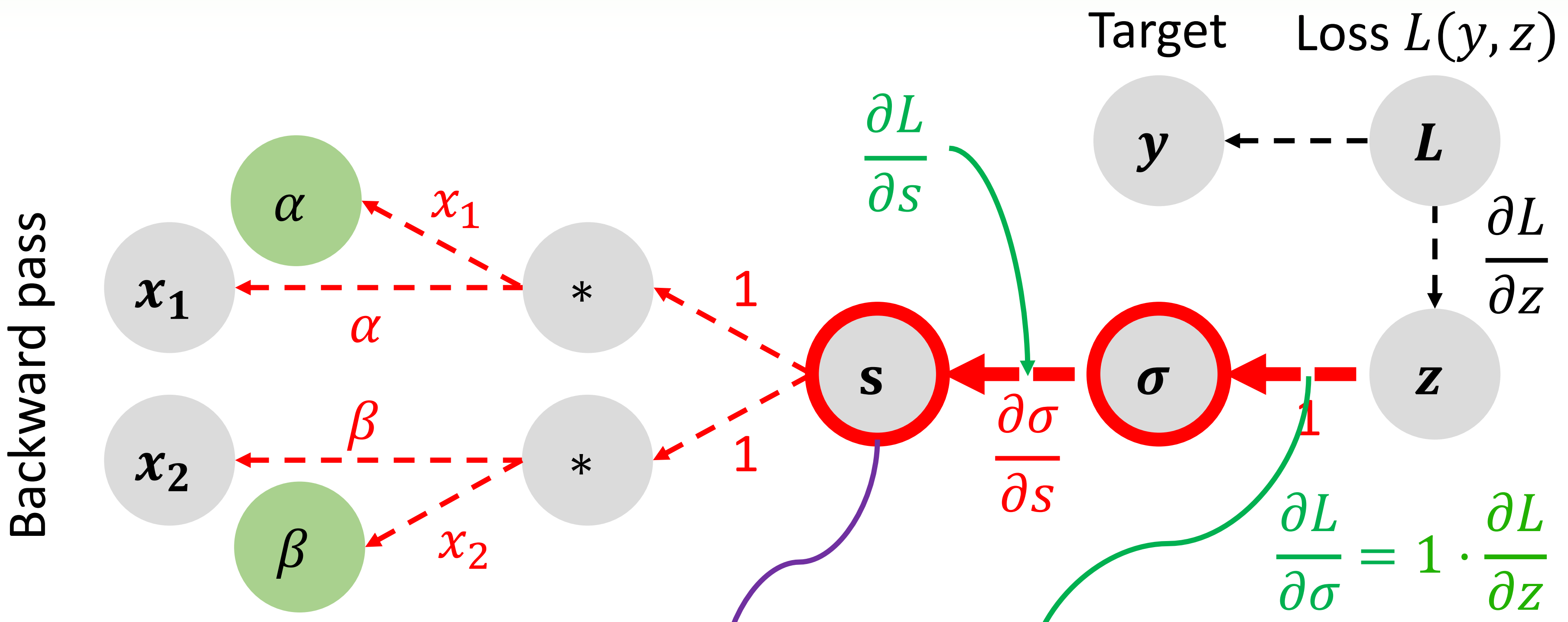
Prediction $z = \sigma(\underbrace{\boldsymbol{\alpha} x_1 + \boldsymbol{\beta} x_2}_{s})$

```
def forward_pass(inputs):
    return 1. / (1 + np.exp(-inputs))
```

Интерфейс обратного прохода



Интерфейс обратного прохода




```
def backward_pass(inputs, incoming_gradient):
    sigmoid = 1. / (1 + np.exp(-inputs))
    return sigmoid * (1 - sigmoid) * incoming_gradient
```

$$\frac{\partial L}{\partial s} = \frac{\partial \sigma}{\partial s} \cdot \frac{\partial L}{\partial \sigma}$$

Полносвязный слой как произведение матриц

- Пример для 2 нейронов с линейной активацией, 3 входами, без свободных членов:

$$(x_1 \quad x_2 \quad x_3) \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = (z_1 \quad z_2)$$

$$z_1 = x_1 w_{1,1} + x_2 w_{2,1} + x_3 w_{3,1}$$
$$z_2 = x_1 w_{1,2} + x_2 w_{2,2} + x_3 w_{3,2}$$

$$xW = z$$

- Быстрые матричные операции на CPU (BLAS) и GPU (cuBLAS).

- Матричные операции в NumPy сильно быстрее циклов в Python.

Обратный проход

- Прямой проход:

$$xW = z \qquad\qquad (x_1 \quad x_2 \quad x_3) \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = (z_1 \quad z_2)$$

- Для обратного прохода нужен $\frac{\partial L}{\partial W}$, где $L(z_1, z_2)$ – скалярные потери.

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial w_{1,1}} & \frac{\partial L}{\partial w_{1,2}} \\ \frac{\partial L}{\partial w_{2,1}} & \frac{\partial L}{\partial w_{2,2}} \\ \frac{\partial L}{\partial w_{3,1}} & \frac{\partial L}{\partial w_{3,2}} \end{bmatrix}$$

Удобно для SGD:

$$W_{new} = W - \gamma \frac{\partial L}{\partial W}$$

Обратный проход

- Прямой проход:

$$xW = z \qquad \begin{pmatrix} x_1 & x_2 & x_3 \end{pmatrix} \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{pmatrix} z_1 & z_2 \end{pmatrix}$$

- Для обратного прохода нужен $\frac{\partial L}{\partial W}$, где $L(z_1, z_2)$ – скалярные потери.

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial w_{1,1}} & \frac{\partial L}{\partial w_{1,2}} \\ \frac{\partial L}{\partial w_{2,1}} & \frac{\partial L}{\partial w_{2,2}} \\ \frac{\partial L}{\partial w_{3,1}} & \frac{\partial L}{\partial w_{3,2}} \end{bmatrix}$$

Применим цепное правило:

$$\frac{\partial L}{\partial w_{i,j}} = \sum_k \frac{\partial L}{\partial z_k} \frac{\partial z_k}{\partial w_{i,j}} = \frac{\partial L}{\partial z_j} \frac{\partial z_j}{\partial w_{i,j}} = \frac{\partial L}{\partial z_j} x_i$$

$$z_j = x_1 w_{1,j} + x_2 w_{2,j} + x_3 w_{3,j}$$

Обратный проход

- Прямой проход:

$$xW = z \qquad\qquad (x_1 \quad x_2 \quad x_3) \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = (z_1 \quad z_2)$$

- Для обратного прохода нужен $\frac{\partial L}{\partial W}$, где $L(z_1, z_2)$ – скалярные потери.

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \frac{\partial L}{\partial w_{1,1}} & \frac{\partial L}{\partial w_{1,2}} \\ \frac{\partial L}{\partial w_{2,1}} & \frac{\partial L}{\partial w_{2,2}} \\ \frac{\partial L}{\partial w_{3,1}} & \frac{\partial L}{\partial w_{3,2}} \end{bmatrix}$$

Перепишем в матричном виде:

$$\frac{\partial L}{\partial w_{i,j}} = \sum_k \frac{\partial L}{\partial z_k} \frac{\partial z_k}{\partial w_{i,j}} = \frac{\partial L}{\partial z_j} \frac{\partial z_j}{\partial w_{i,j}} = \frac{\partial L}{\partial z_j} x_i$$

$$\frac{\partial L}{\partial z} = \left(\frac{\partial L}{\partial z_1} \quad \frac{\partial L}{\partial z_2} \right)$$ градиент

$$\frac{\partial L}{\partial W} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \cdot \left(\frac{\partial L}{\partial z_1} \quad \frac{\partial L}{\partial z_2} \right) = x^T \frac{\partial L}{\partial z}$$

Прямой проход для мини-батча

Батч из 2: $\begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{pmatrix} \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \end{pmatrix}$

Матричный вид: $XW = Z$

1 нейрон для 2 примера: $z_{\mathbf{2},1} = x_{\mathbf{2},1} w_{1,1} + x_{\mathbf{2},2} w_{2,1} + x_{\mathbf{2},3} w_{3,1}$

Обратный проход для мини-батча

Батч из 2: $$\begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{pmatrix} \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \end{pmatrix}$$

SGD шаг: $$\frac{\partial L_b}{\partial W} = \frac{\partial L\left(z_{\mathbf{1},1}, z_{\mathbf{1},2}\right)}{\partial W} + \frac{\partial L\left(z_{\mathbf{2},1}, z_{\mathbf{2},2}\right)}{\partial W}$$

Для одного семпла: $$\frac{\partial L}{\partial w_{i,j}} = \frac{\partial L}{\partial z_j}\frac{\partial z_j}{\partial w_{i,j}} = \frac{\partial L}{\partial z_j} x_i$$ *это уже знаем*

Для 2 семплов: $$\frac{\partial L_b}{\partial w_{i,j}} = \frac{\partial L}{\partial z_{\mathbf{1},j}} x_{\mathbf{1},i} + \frac{\partial L}{\partial z_{\mathbf{2},j}} x_{\mathbf{2},i}$$

Обратный проход для мини-батча

Батч из 2: $\begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{pmatrix} \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \end{pmatrix}$

SGD шаг: $\dfrac{\partial L_b}{\partial W} = \dfrac{\partial L\left(z_{\mathbf{1},1}, z_{\mathbf{1},2}\right)}{\partial W} + \dfrac{\partial L\left(z_{\mathbf{2},1}, z_{\mathbf{2},2}\right)}{\partial W}$

Для 2 семплов: $\dfrac{\partial L_b}{\partial w_{i,j}} = \dfrac{\partial L}{\partial z_{1,j}} x_{1,i} + \dfrac{\partial L}{\partial z_{2,j}} x_{2,i}$

$$\frac{\partial L_b}{\partial W} = X^T \frac{\partial L}{\partial Z} \quad \Bigg| \quad X^T = \begin{pmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \\ x_{1,3} & x_{2,3} \end{pmatrix} \quad \Bigg| \quad \frac{\partial L}{\partial Z} = \begin{pmatrix} \dfrac{\partial L}{\partial z_{1,1}} & \dfrac{\partial L}{\partial z_{1,2}} \\ \dfrac{\partial L}{\partial z_{2,1}} & \dfrac{\partial L}{\partial z_{2,2}} \end{pmatrix}$$

Обратный проход для мини-батча

Батч из 2: $\begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{pmatrix} \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \end{pmatrix}$

SGD шаг: $\frac{\partial L_b}{\partial W} = \frac{\partial L\left(z_{\mathbf{1},1}, z_{\mathbf{1},2}\right)}{\partial W} + \frac{\partial L\left(z_{\mathbf{2},1}, z_{\mathbf{2},2}\right)}{\partial W}$

Для 2 семплов: $\frac{\partial L_b}{\partial \boldsymbol{w}_{\mathbf{3,2}}} = \frac{\partial L}{\partial z_{1,2}} x_{1,3} + \frac{\partial L}{\partial z_{2,2}} x_{2,3}$ *Проверка!*

$$\frac{\partial L_b}{\partial W} = X^T \frac{\partial L}{\partial Z} \quad \Bigg| \quad X^T = \begin{pmatrix} x_{1,1} & x_{2,1} \\ x_{1,2} & x_{2,2} \\ \boldsymbol{x}_{\mathbf{1,3}} & \boldsymbol{x}_{\mathbf{2,3}} \end{pmatrix} \quad \Bigg| \quad \frac{\partial L}{\partial Z} = \begin{pmatrix} \frac{\partial L}{\partial z_{1,1}} & \boldsymbol{\frac{\partial L}{\partial z_{1,2}}} \\ \frac{\partial L}{\partial z_{2,1}} & \boldsymbol{\frac{\partial L}{\partial z_{2,2}}} \end{pmatrix}$$

Обратный проход для X (для глубоких слоев)

Батч из 2:
$$\begin{pmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \end{pmatrix} \cdot \begin{bmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{bmatrix} = \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \end{pmatrix}$$

SGD шаг:
$$\frac{\partial L_b}{\partial X} = \frac{\partial L(z_{\mathbf{1},1}, z_{\mathbf{1},2})}{\partial X} + \frac{\partial L(z_{\mathbf{2},1}, z_{\mathbf{2},2})}{\partial X}$$

Для 1 семпла: *Применим цепное правило!*

Примеры независимы!
$$\frac{\partial L(z_{\boldsymbol{i},1}, z_{\boldsymbol{i},2})}{\partial x_{\boldsymbol{i},j}} = \sum_k \frac{\partial L}{\partial z_{i,k}} \frac{\partial z_{i,k}}{\partial x_{i,j}} = \sum_k \frac{\partial L}{\partial z_{i,k}} w_{j,k} = \sum_k \frac{\partial L}{\partial z_{i,k}} w_{k,j}^T$$

Для 2 семплов:
$$\frac{\partial L_b}{\partial X} = \frac{\partial L}{\partial Z} W^T$$

дает одну ненулевую строку!

Быстрая реализация в NumPy

- Прямой ход:

```
def forward_pass(X, W):
    return X.dot(W)
```

$$XW = Z$$

- Обратный ход:

```
def backward_pass(X, W, dZ):
    dX = dZ.dot(W.T)
    dW = X.T.dot(dZ)
    return dX, dW
```

$$\frac{\partial L_b}{\partial X} = \frac{\partial L}{\partial Z} W^T$$

$$\frac{\partial L_b}{\partial W} = X^T \frac{\partial L}{\partial Z}$$

- Еще одна причина иметь $\frac{\partial L}{\partial Z}$ в интерфейсе обратного шага:
    - Иначе пришлось бы считать $\frac{\partial Z}{\partial X}$ и $\frac{\partial Z}{\partial W}$, а это тензоры (многомерные массивы)!

Резюме

- **Плюсы:**
  - Универсальные аппроксиматоры (приближают сложные функции)
  - Сложные композиции простых функций (легко дифференцировать)

- **Минусы:**
  - Архитектуру надо подбирать руками
  - Сильное переобучение (нужна регуляризация)

Ссылки

- https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/
- http://www.machinelearning.ru/wiki/images/c/c2/Voron-ML-NeuralNets-slides.pdf