Южный федеральный университет

@SINECOR

# Обратное распространение ошибки и SGD



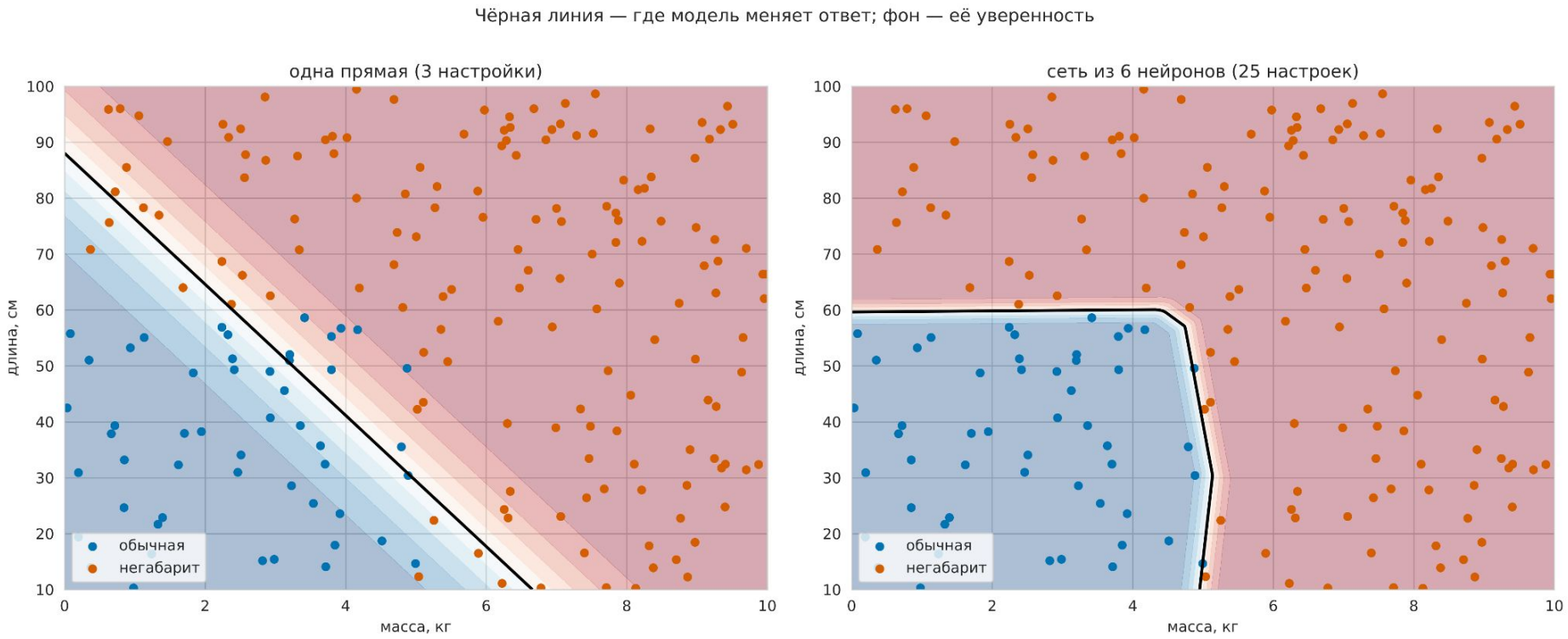
Сергей Николенко
2 ноября 2026 г.

**Технологии и фронтиры науки о данных**

# Что такое градиентный спуск?

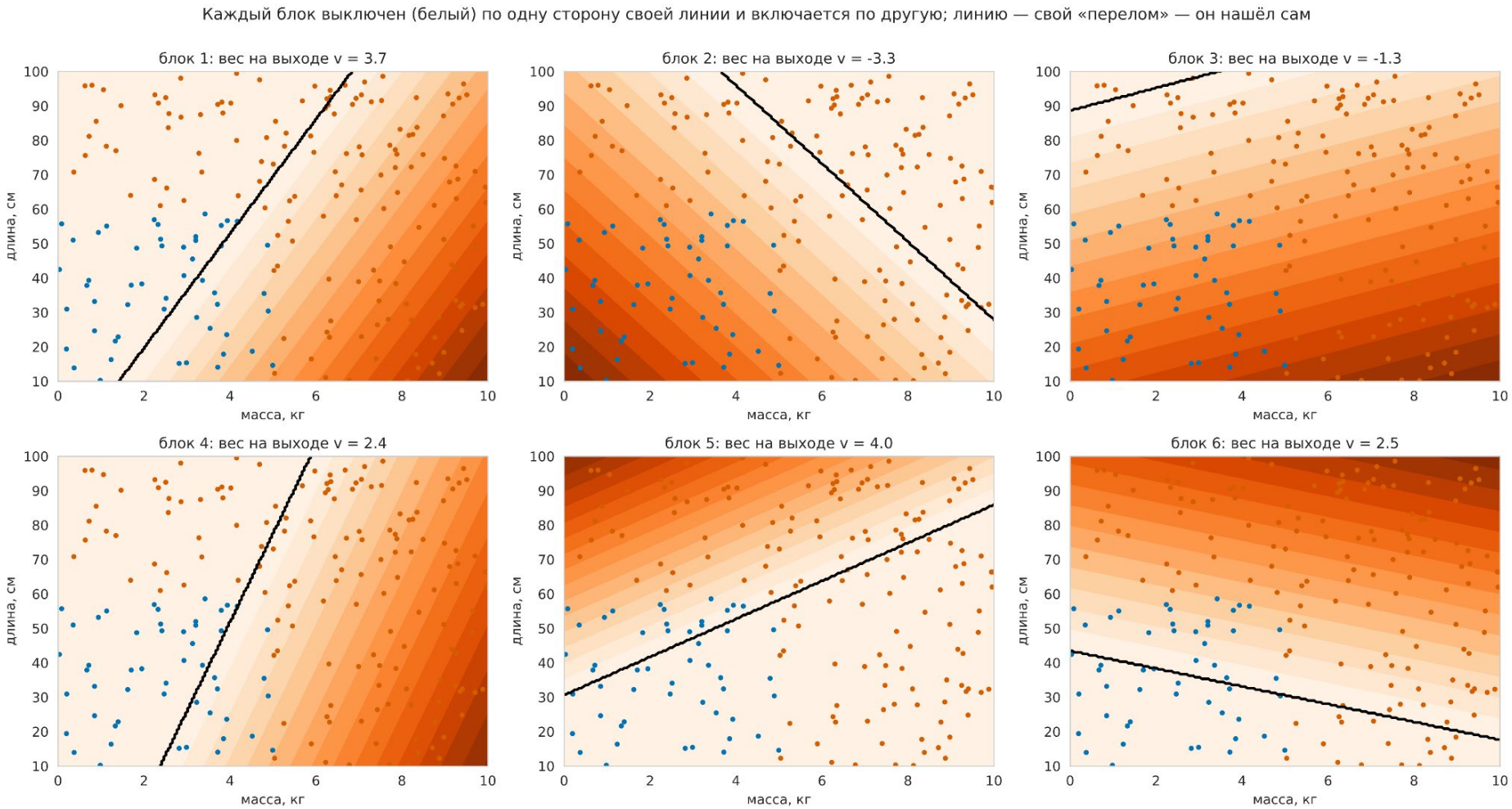


# Классификация




- Вспомним наш пример с классификацией
- Здесь, например, просто надо вырезать прямоугольничек
- Прямая не может этого сделать, а нейросеть со скрытым слоем легко

# Классификация

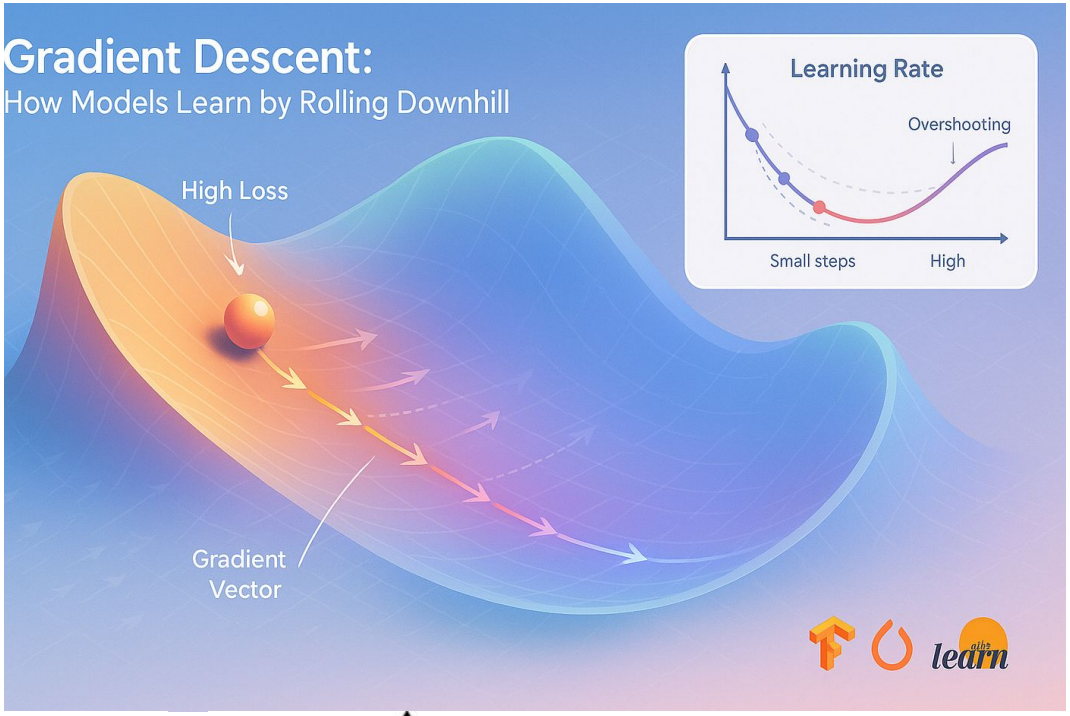


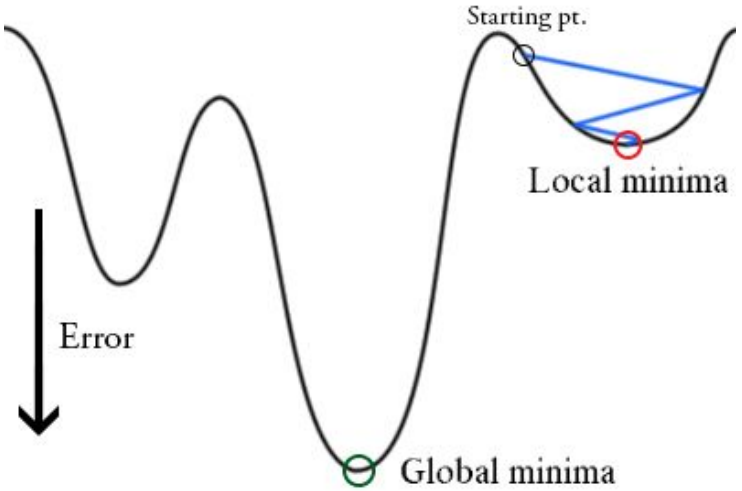

- ...её нейроны, правда, не так легко интерпретировать

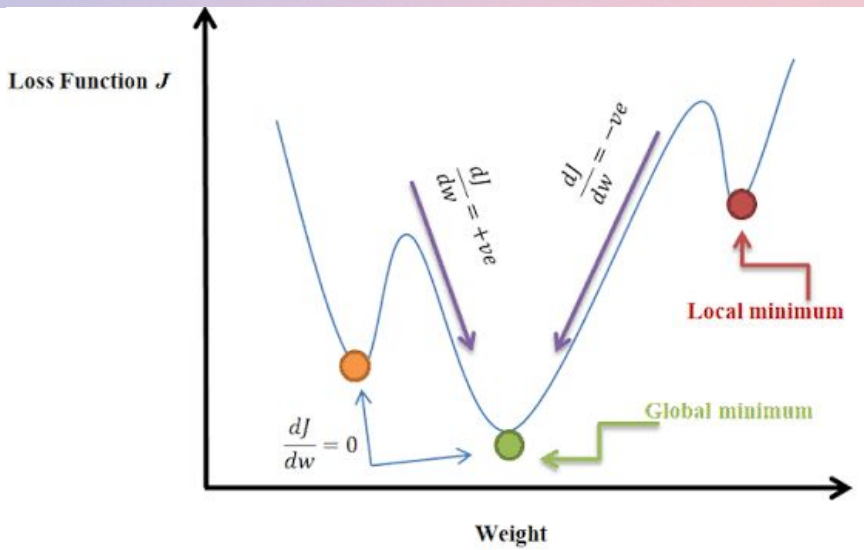
# Градиентный спуск

- Самый главный метод оптимизации: нужно спускаться в сторону, обратную градиенту

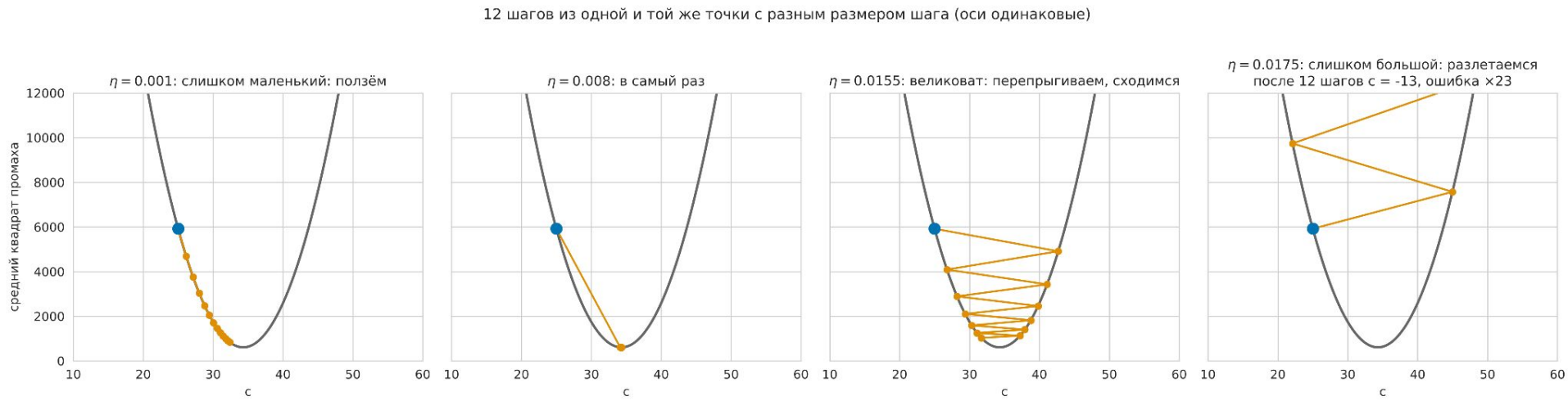
$$\mathbf{x}_k = \mathbf{x}_{k-1} - \alpha \nabla f(\mathbf{x}_k).$$



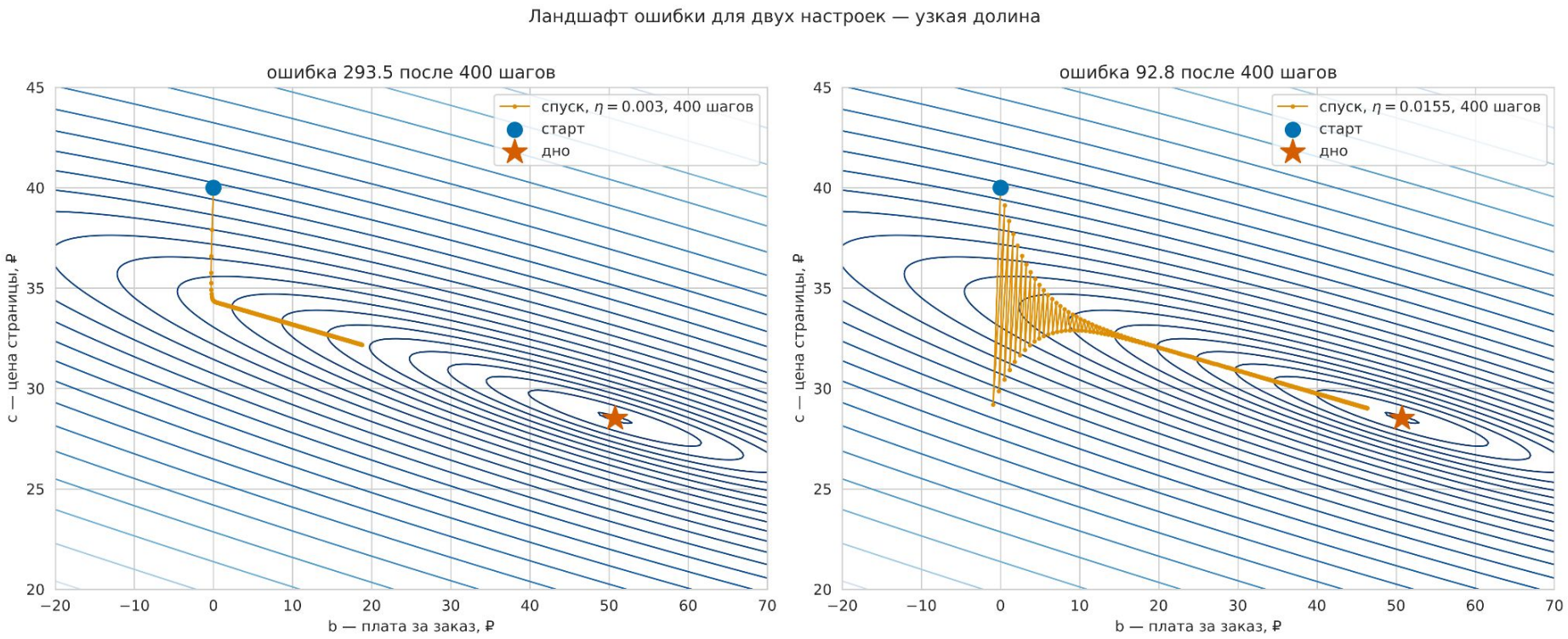



# Скорость обучения

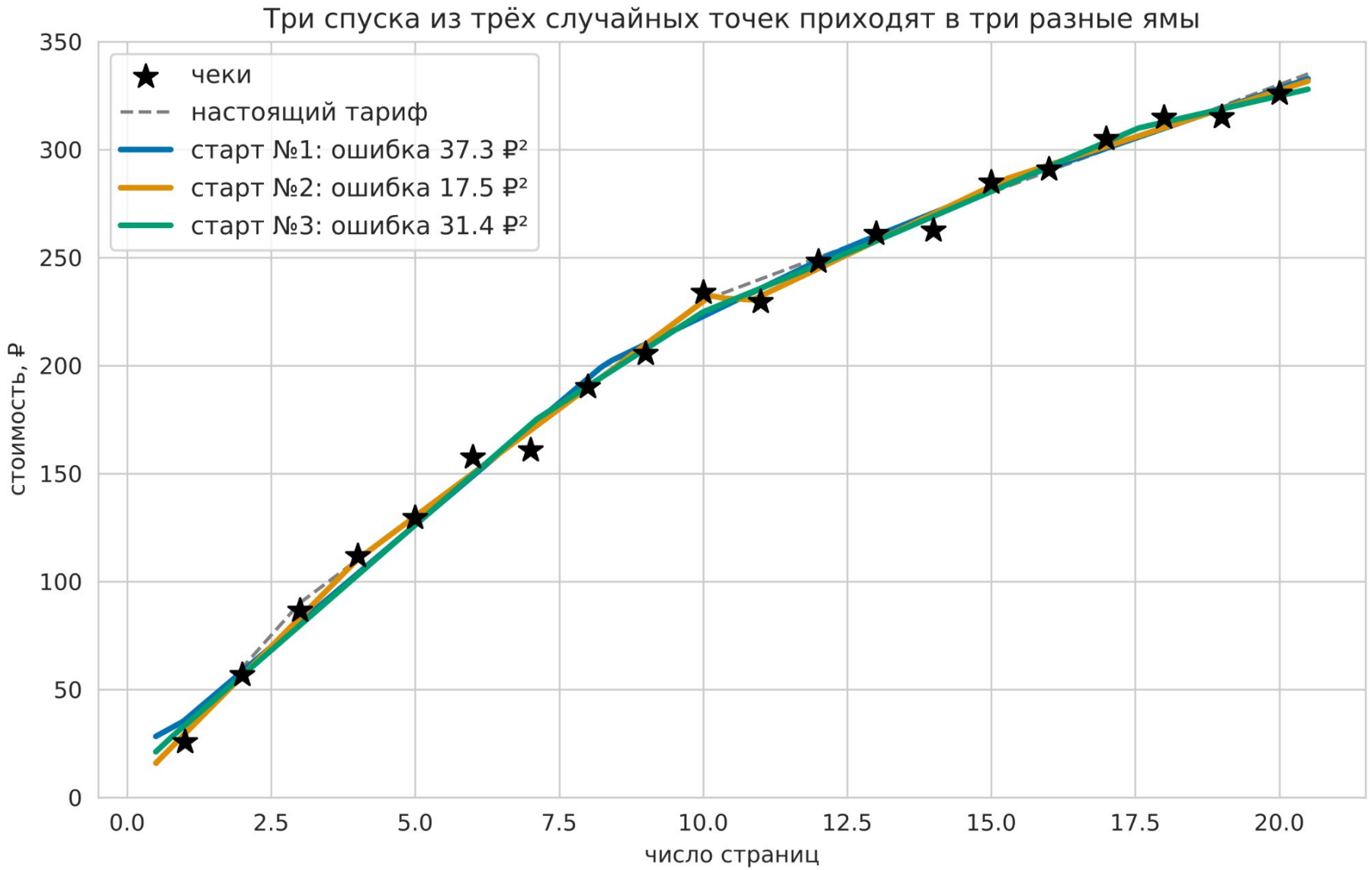



- В таких методах оптимизации (локальных, на градиентном спуске) многое зависит от скорости ошибки

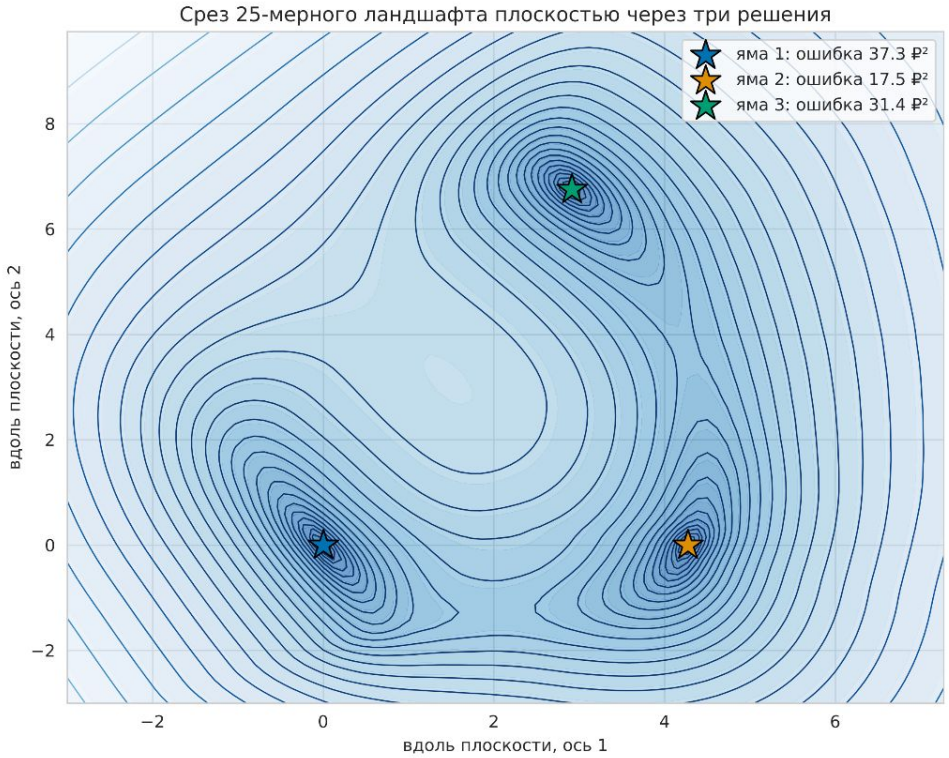
# Ландшафт функции ошибки

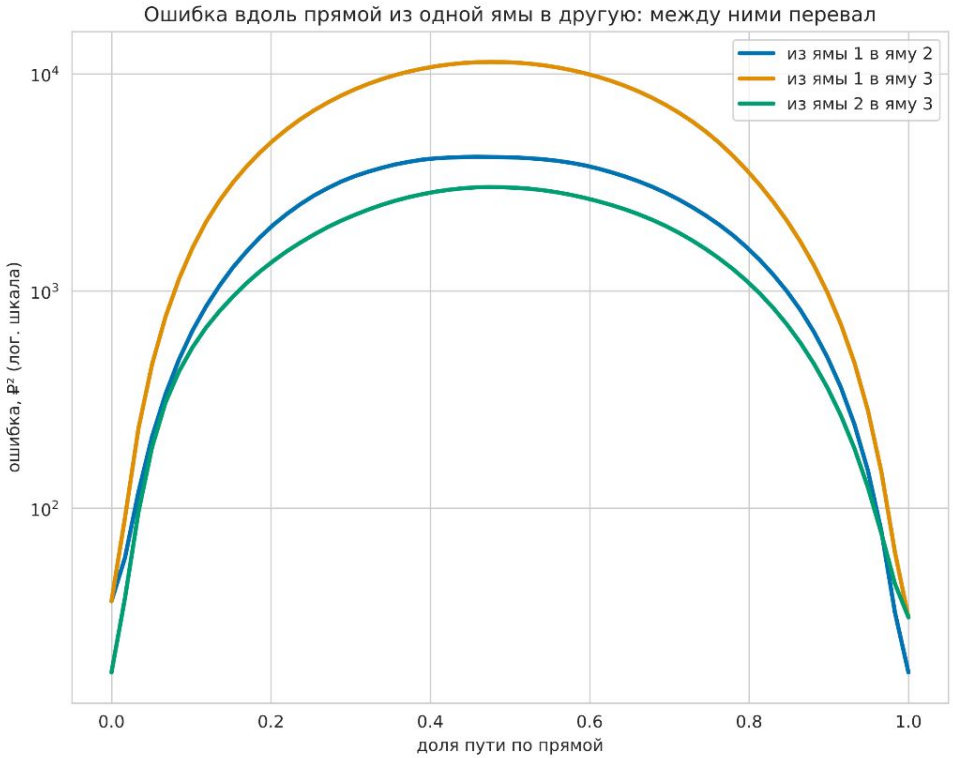


# Разные “ямы” в функции ошибки



# Разные “ямы” в функции ошибки

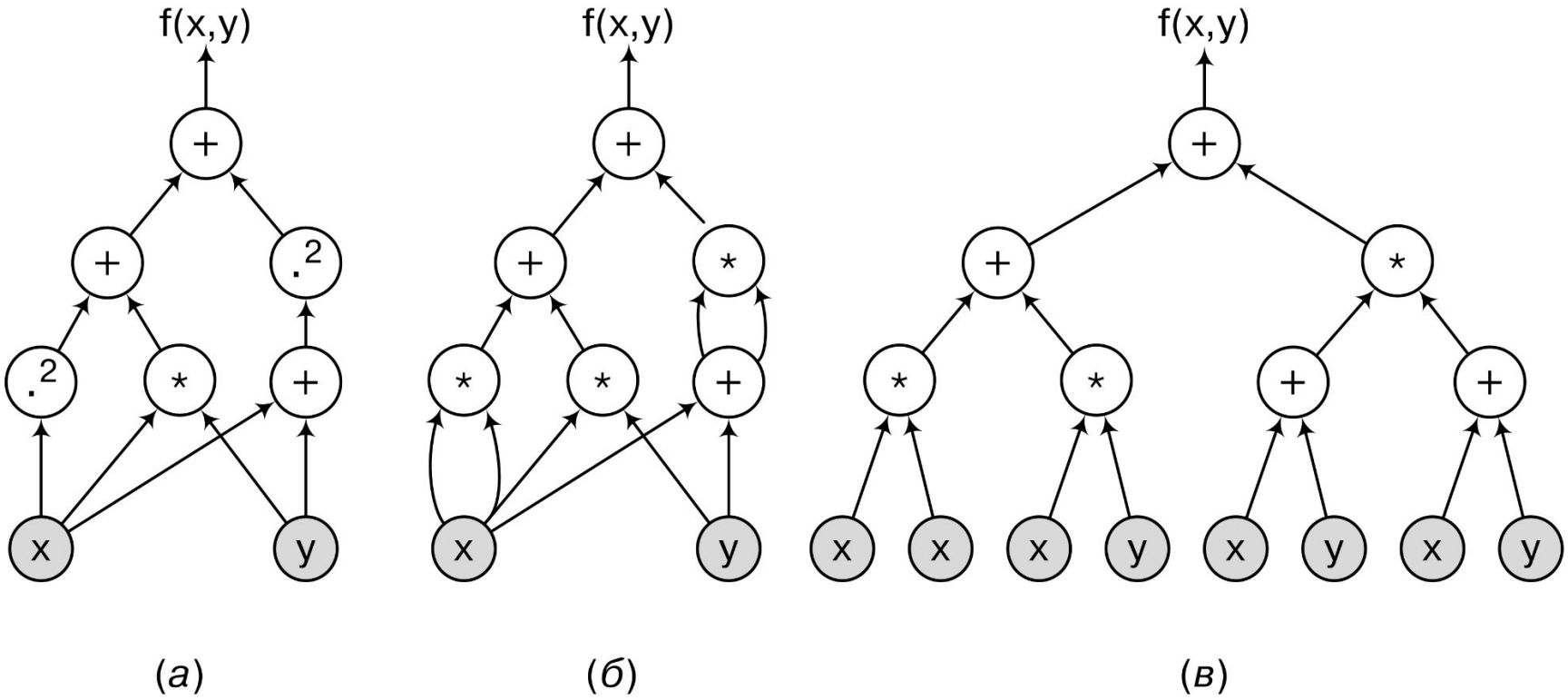



- Из одной ямы в другую можно прийти без больших подъёмов
- Но этот путь не прямой!

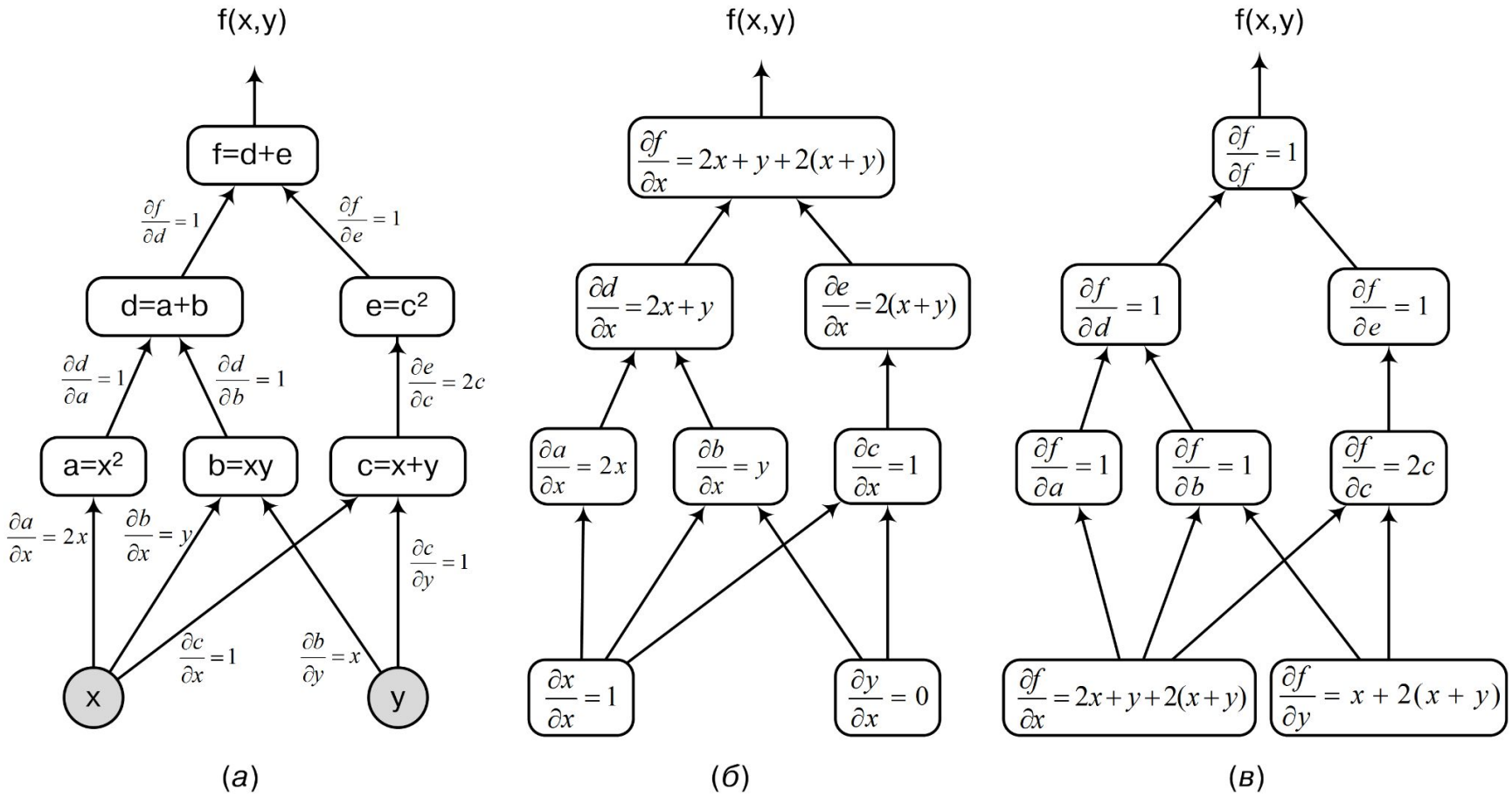
# Обратное распространение ошибки

- Но как подсчитать все эти производные?..
- Рассмотрим пример:



# Обратное распространение ошибки

- Но как подсчитать все эти производные?..
- Рассмотрим пример:

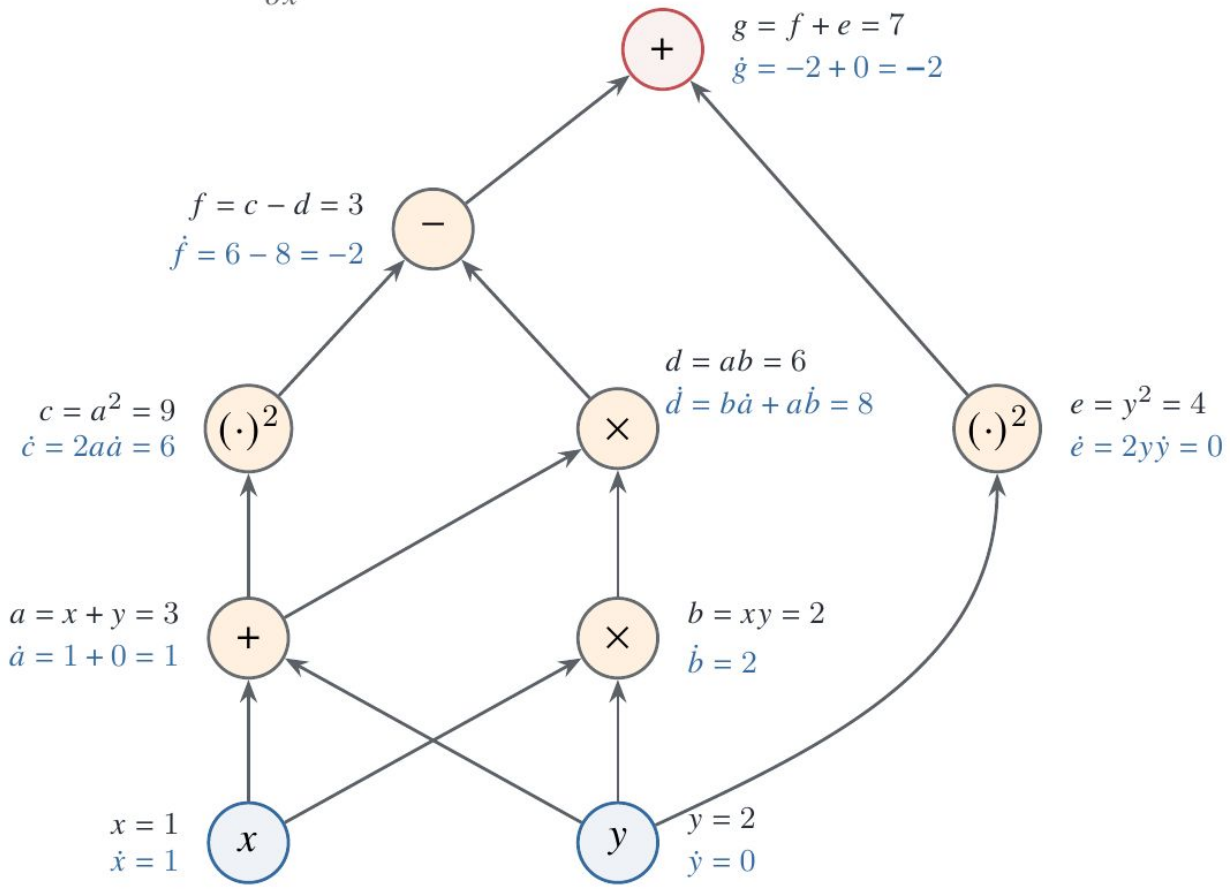


# Наш пример

$$g(x, y) = (x + y)^2 - xy(x + y) + y^2$$

## fprop: прямой проход

$x = 1, \quad y = 2, \qquad \dot{v} = \dfrac{\partial v}{\partial x}$




Вдоль стрелок переносим значение и производную по $x$.

$$\dot{v} = \sum_{u \in \text{parents}(v)} \frac{\partial v}{\partial u} \dot{u}.$$

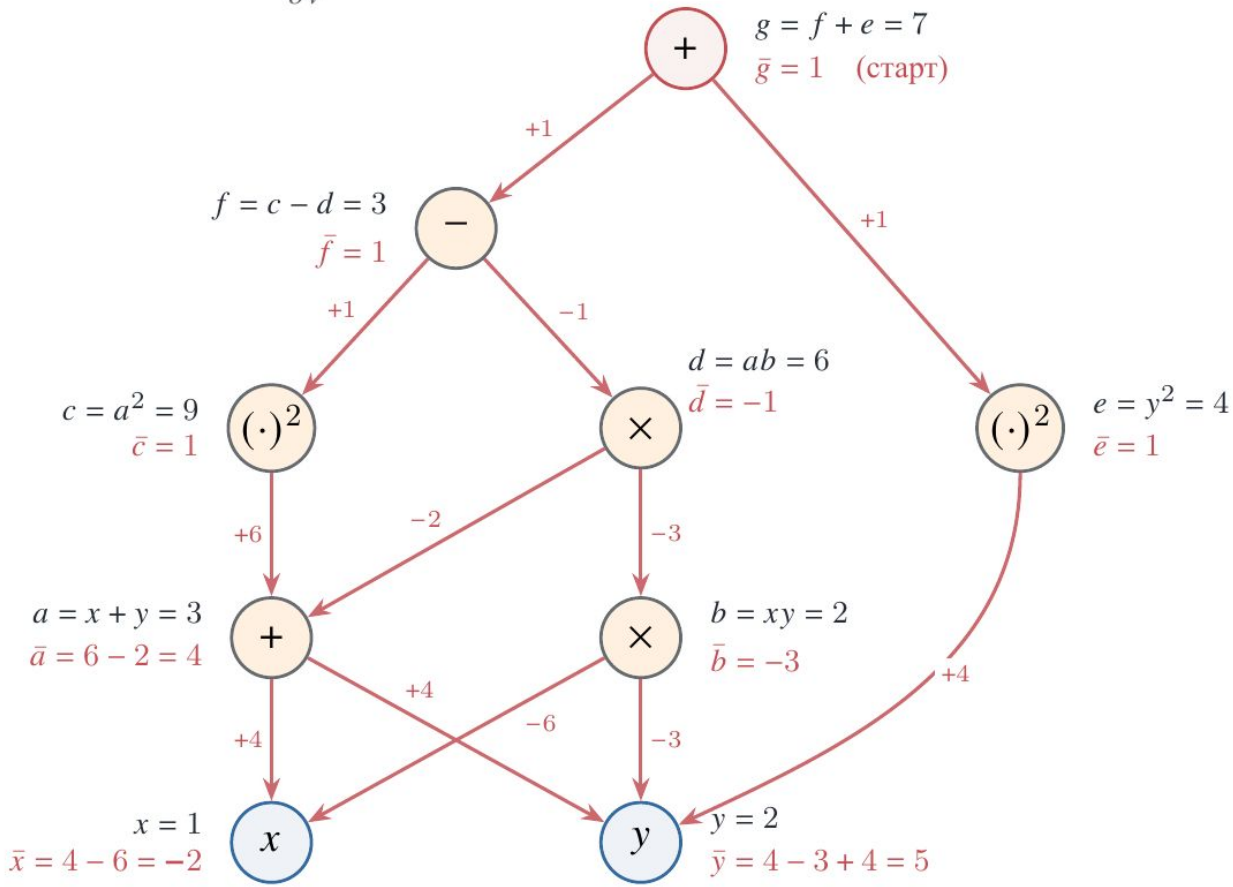
Один прямой проход с $\dot{x} = 1,\ \dot{y} = 0$ даёт $\partial g / \partial x$.

# Наш пример

$$g(x, y) = (x + y)^2 - xy(x + y) + y^2$$

## bprop: обратный проход

$x = 1, \quad y = 2, \qquad \bar{v} = \dfrac{\partial g}{\partial v}$



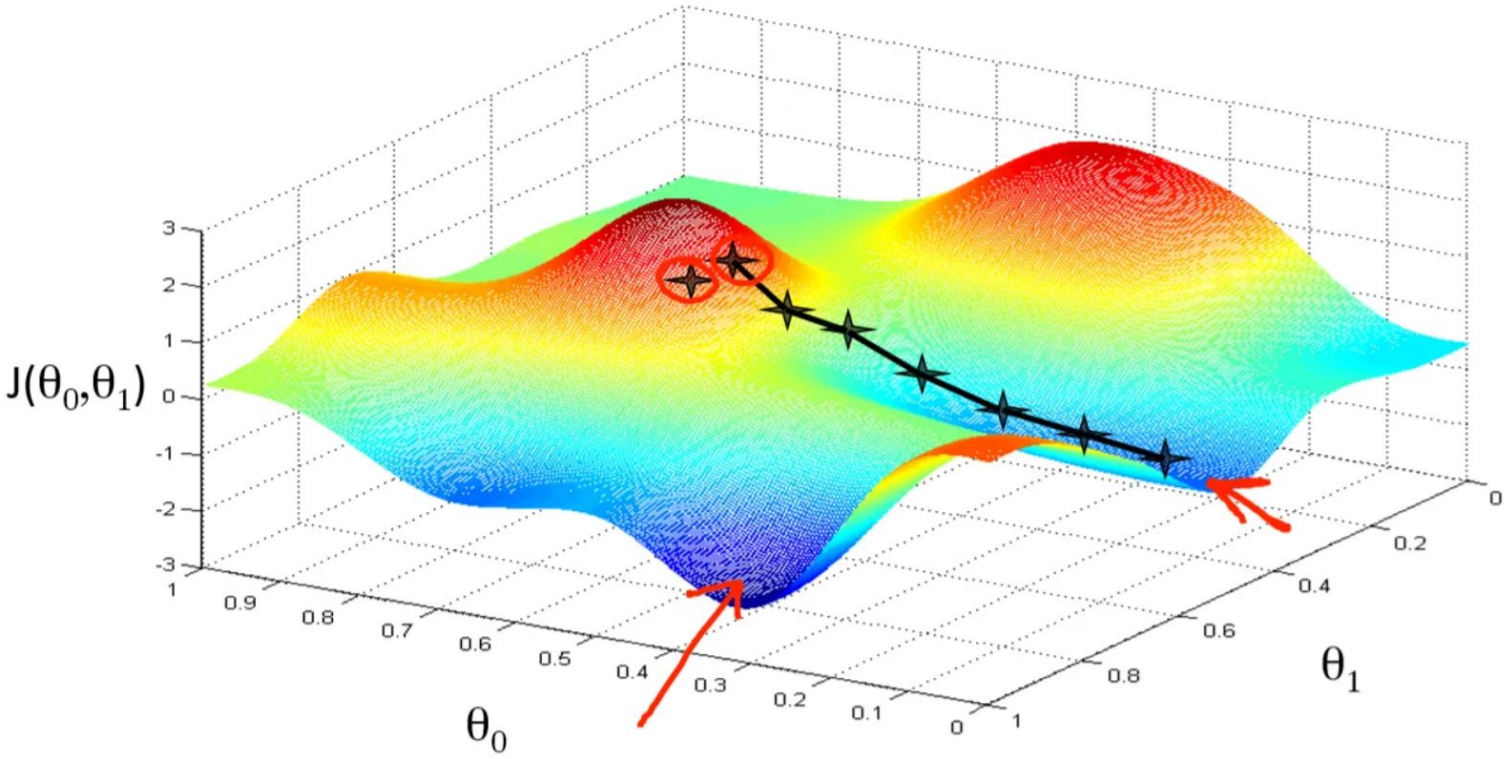

Красные числа на стрелках — вклады; в общем узле складываем их.

$\bar{u} \mathrel{+}= \bar{v}\,\dfrac{\partial v}{\partial u}.$

После вычисления значений один обратный проход даёт $\nabla g(1, 2) = (-2, 5)$.

# Градиентный спуск

- А потом происходит градиентный спуск. Точнее...

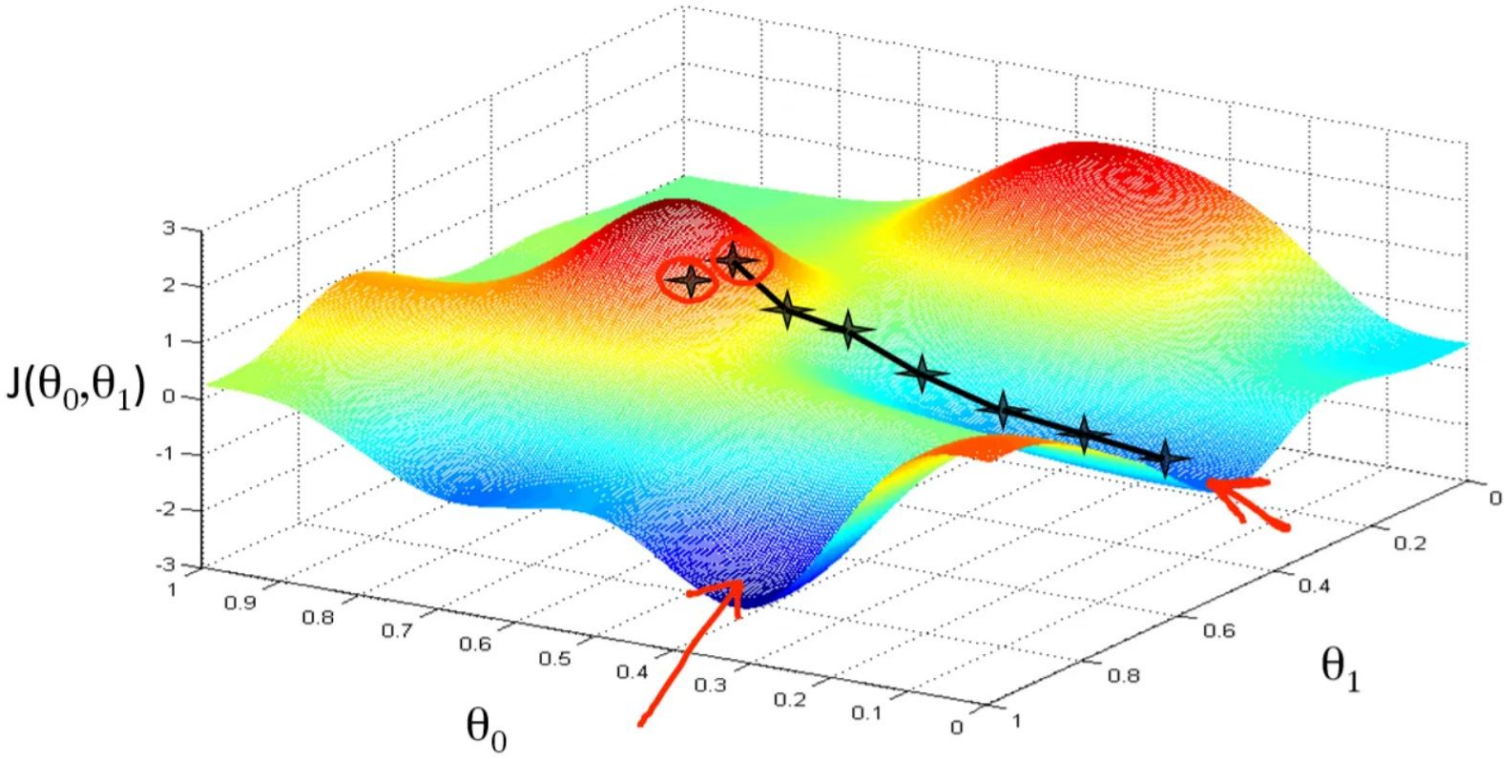


# Стохастический градиентный спуск



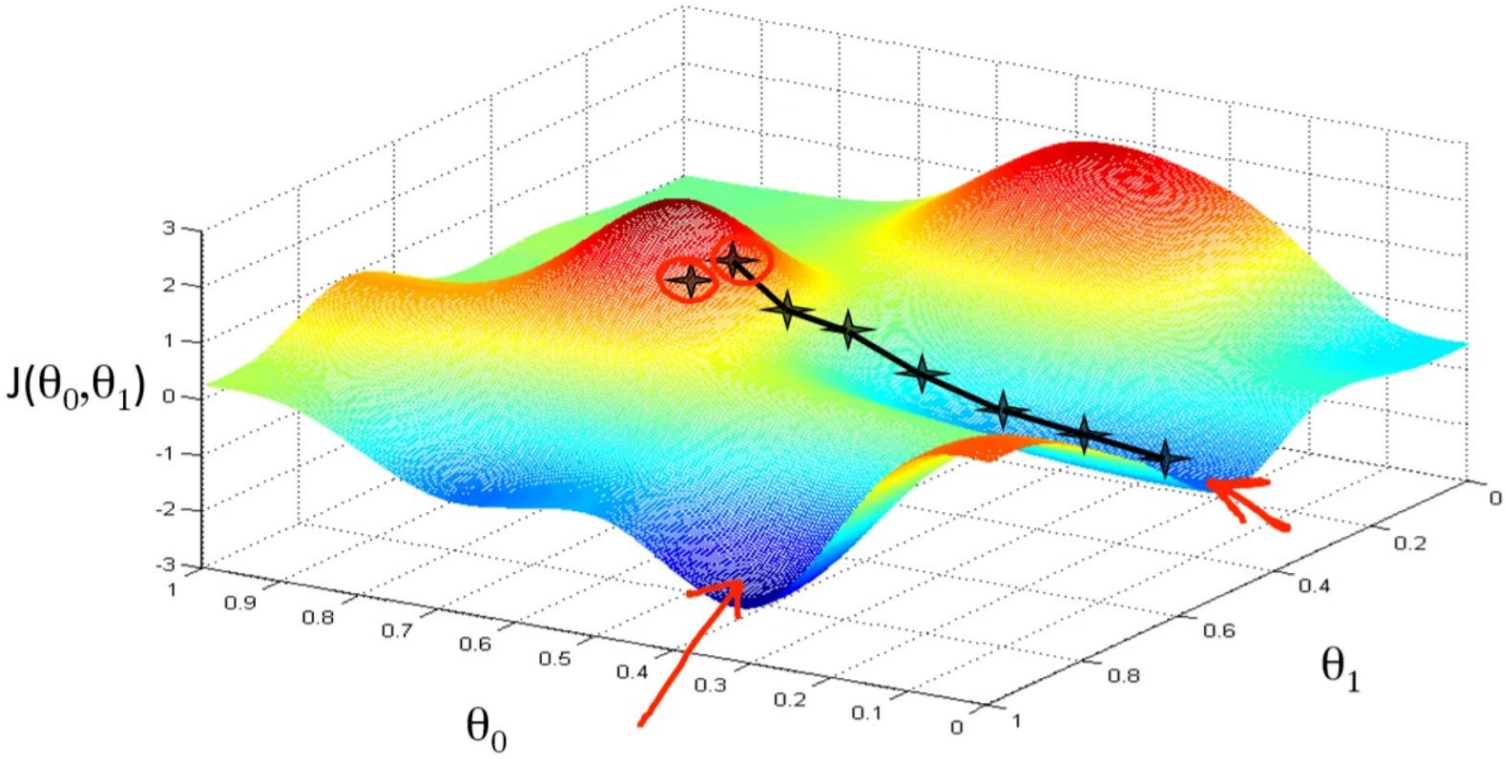
# Градиентный спуск

- Мы уже обсудили, как устроены нейросети, как происходит обратное распространение ошибки, то есть как можно считать градиенты
- А потом происходит градиентный спуск, о котором мы уже говорили. Точнее...



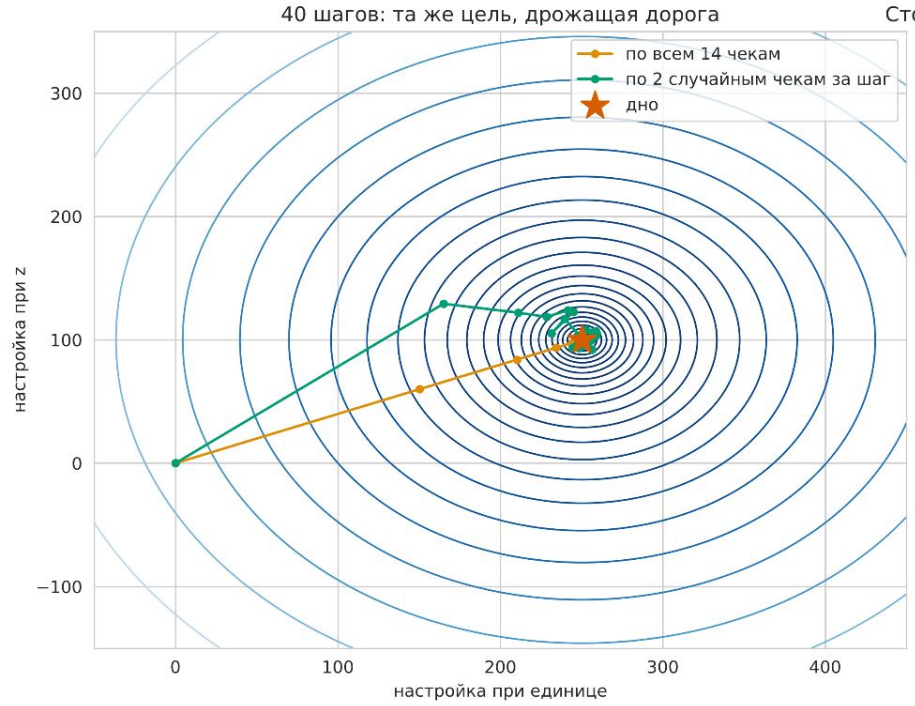
# Градиентный спуск

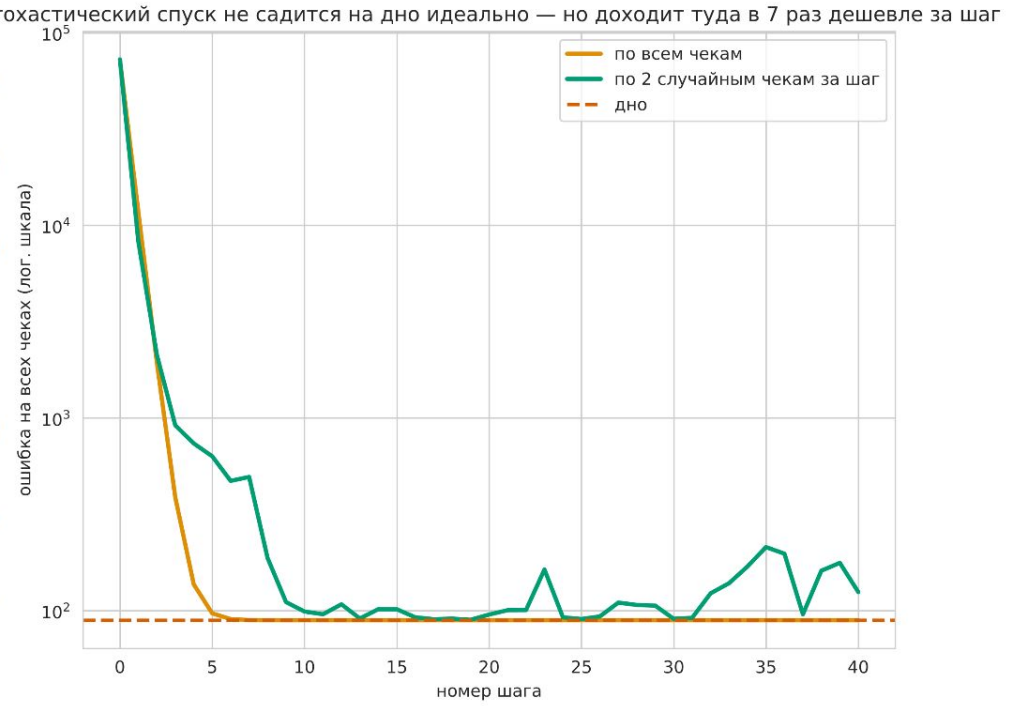
- Мы уже обсудили, как устроены нейросети, как происходит обратное распространение ошибки, то есть как можно считать градиенты
- А потом происходит градиентный спуск, о котором мы уже говорили. Точнее, *стохастический градиентный спуск* (stochastic gradient descent, SGD), потому что вычислить всю функцию ошибки было бы слишком дорого



# Градиентный спуск

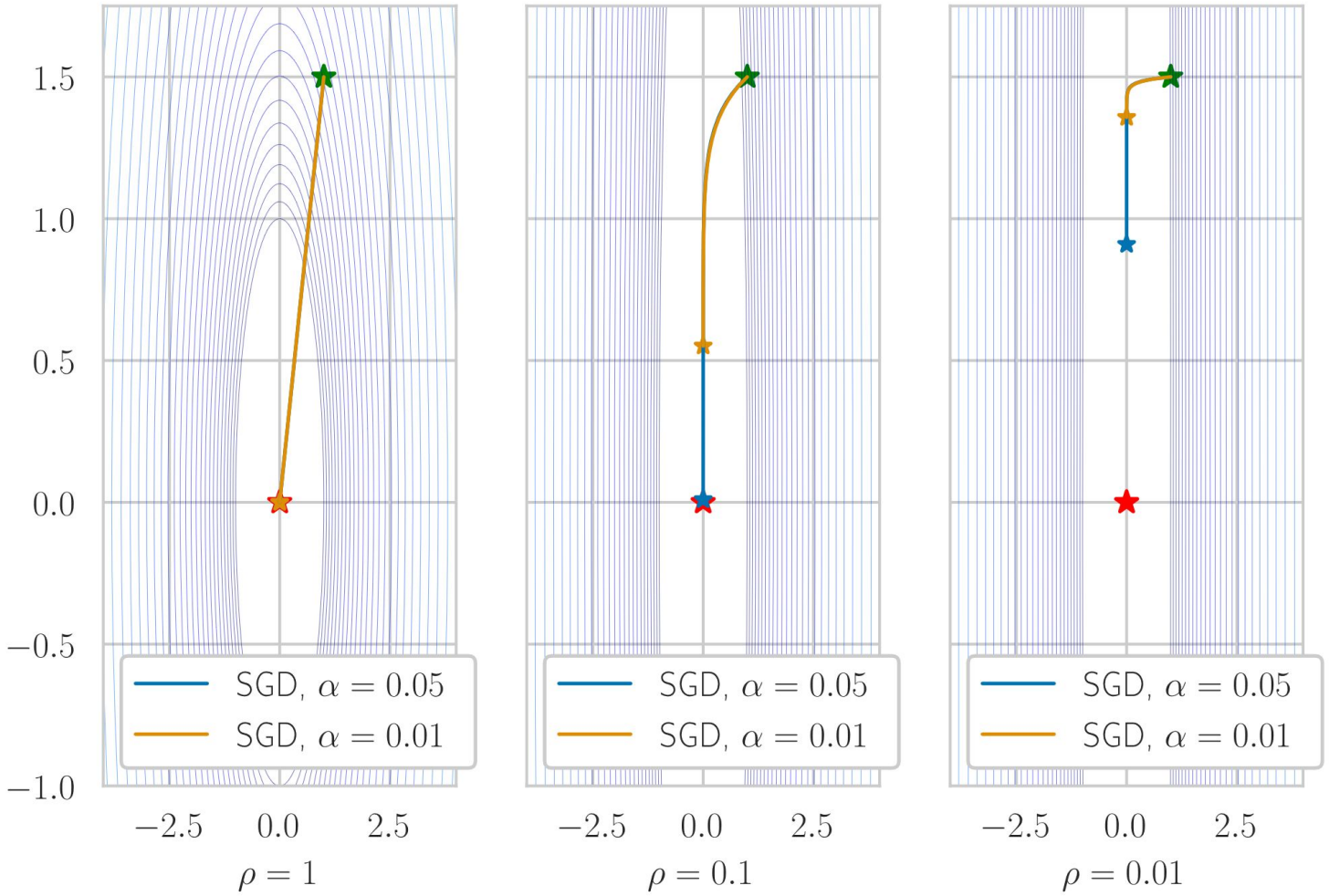
- Даже в наших примерах стохастически было бы быстрее
- Но не сошлось бы окончательно, что логично:



# Градиентный спуск

- Главные проблемы начинаются в высоких размерностях, при разном масштабе переменных:



Южный федеральный университет

# Спасибо за внимание!

@SINECOR

