Южный федеральный университет

@SINECOR

Градиентный спуск: как обучаются нейронные сети



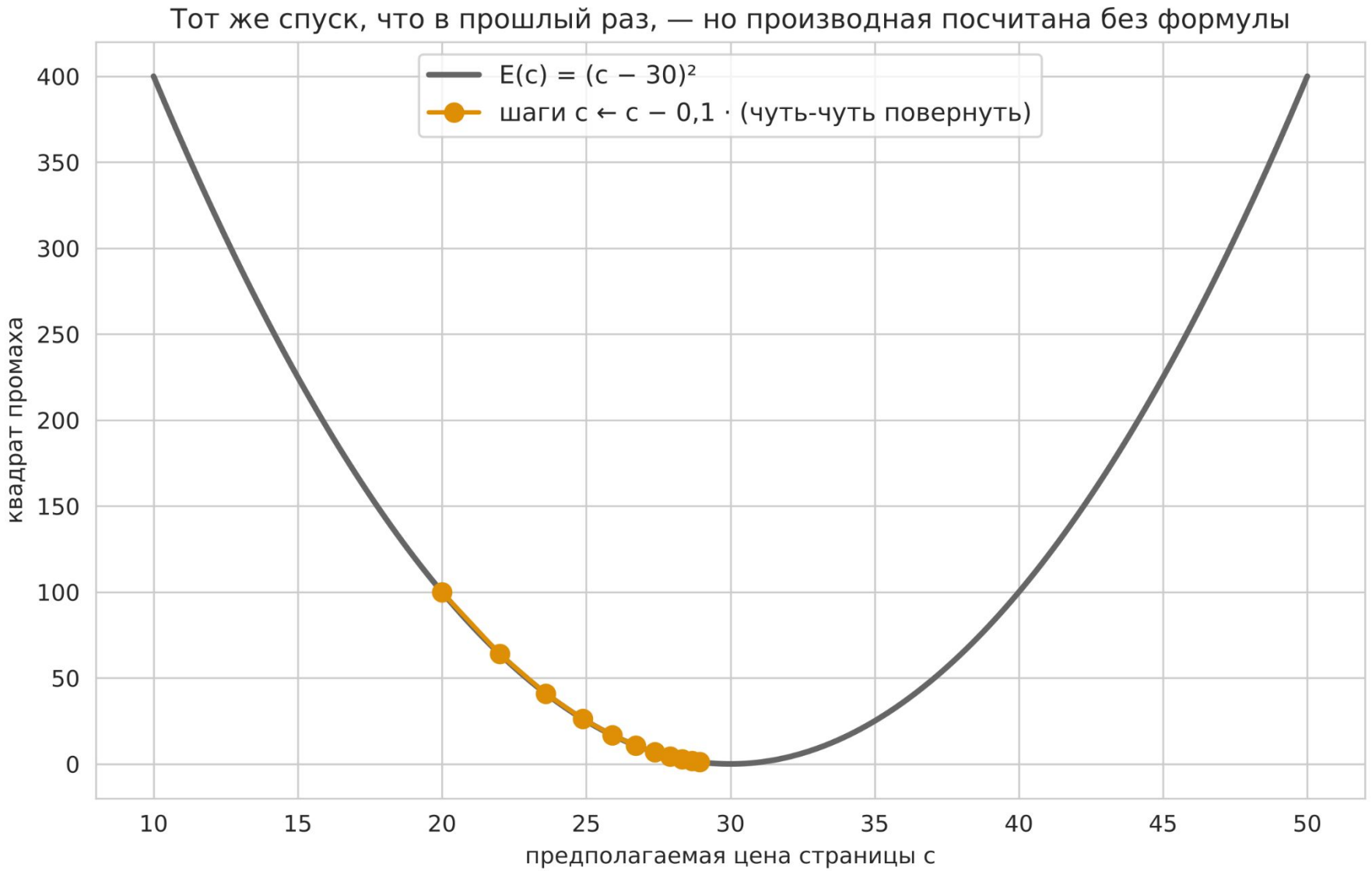
Сергей Николенко
12 октября 2026 г.

Технологии и фронтиры науки о данных

Что такое градиент?

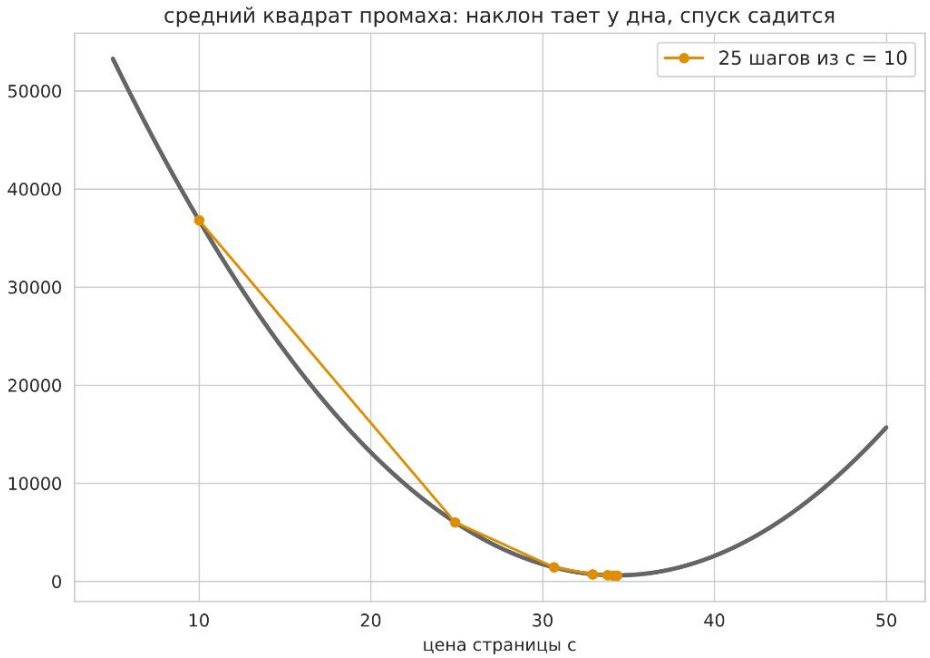


Вернёмся к примеру из прошлой лекции



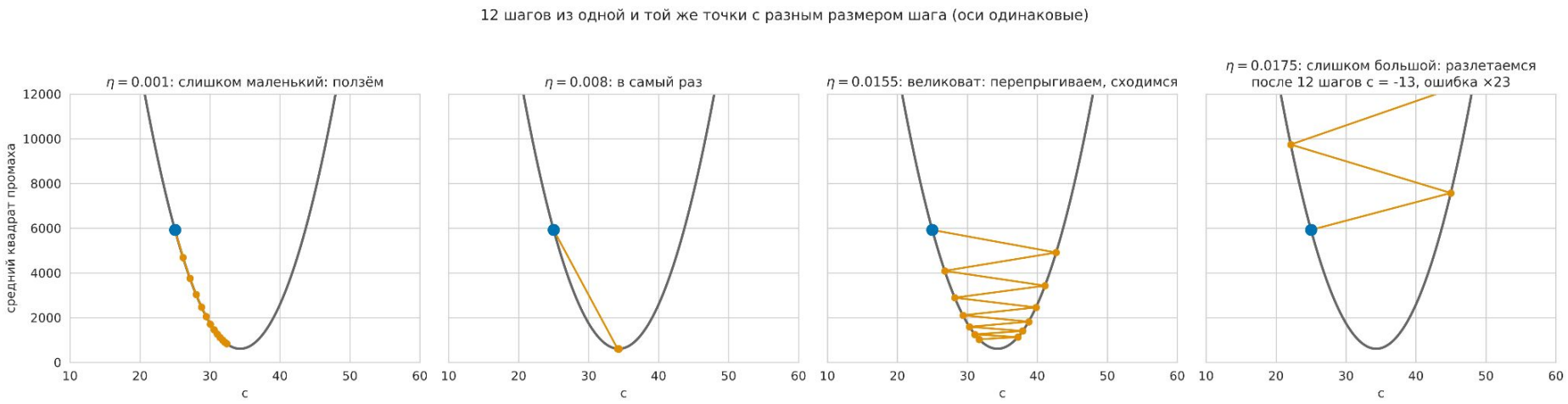
Модуль или квадрат?





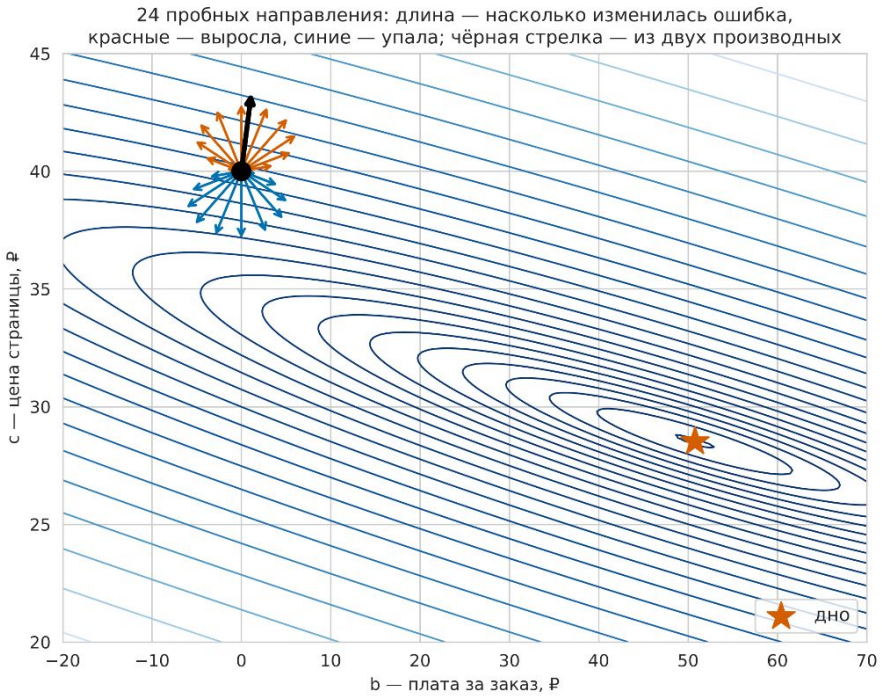
- При модуле отклонения производная всё время постоянная
- А при квадрате отклонения она убывает до нуля в точке минимума
- Так что для нашего локального поиска вроде бы однозначно лучше квадрат (и пока будет он); впрочем, потом это будет не совсем так...

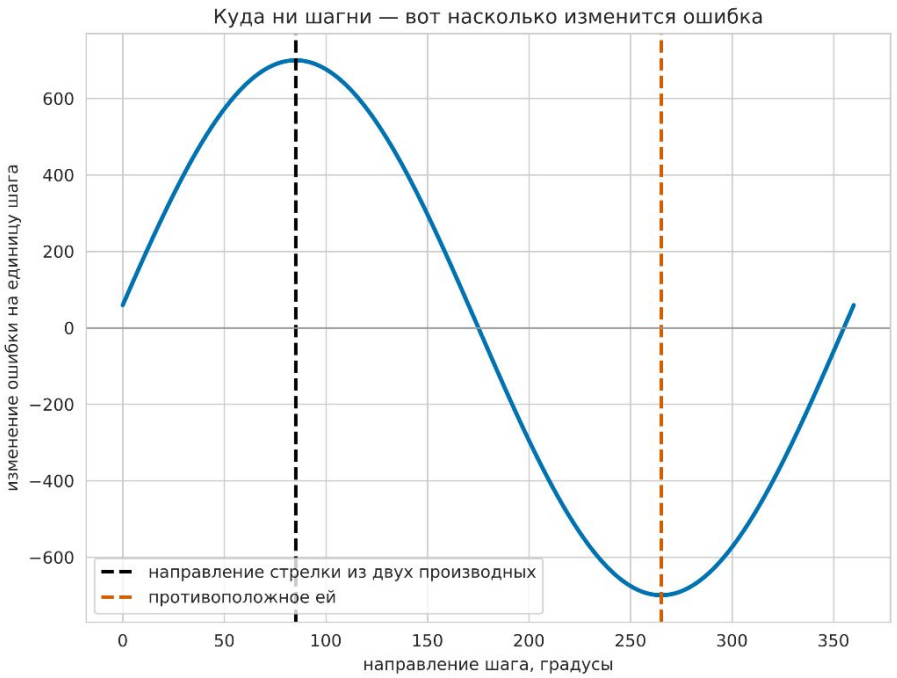
Какую брать скорость обучения?



- Скорость обучения определяет поведение оптимизатора
- Слишком мало – не успеем дойти, слишком много – будем долго перепрыгивать или вообще разойдёмся

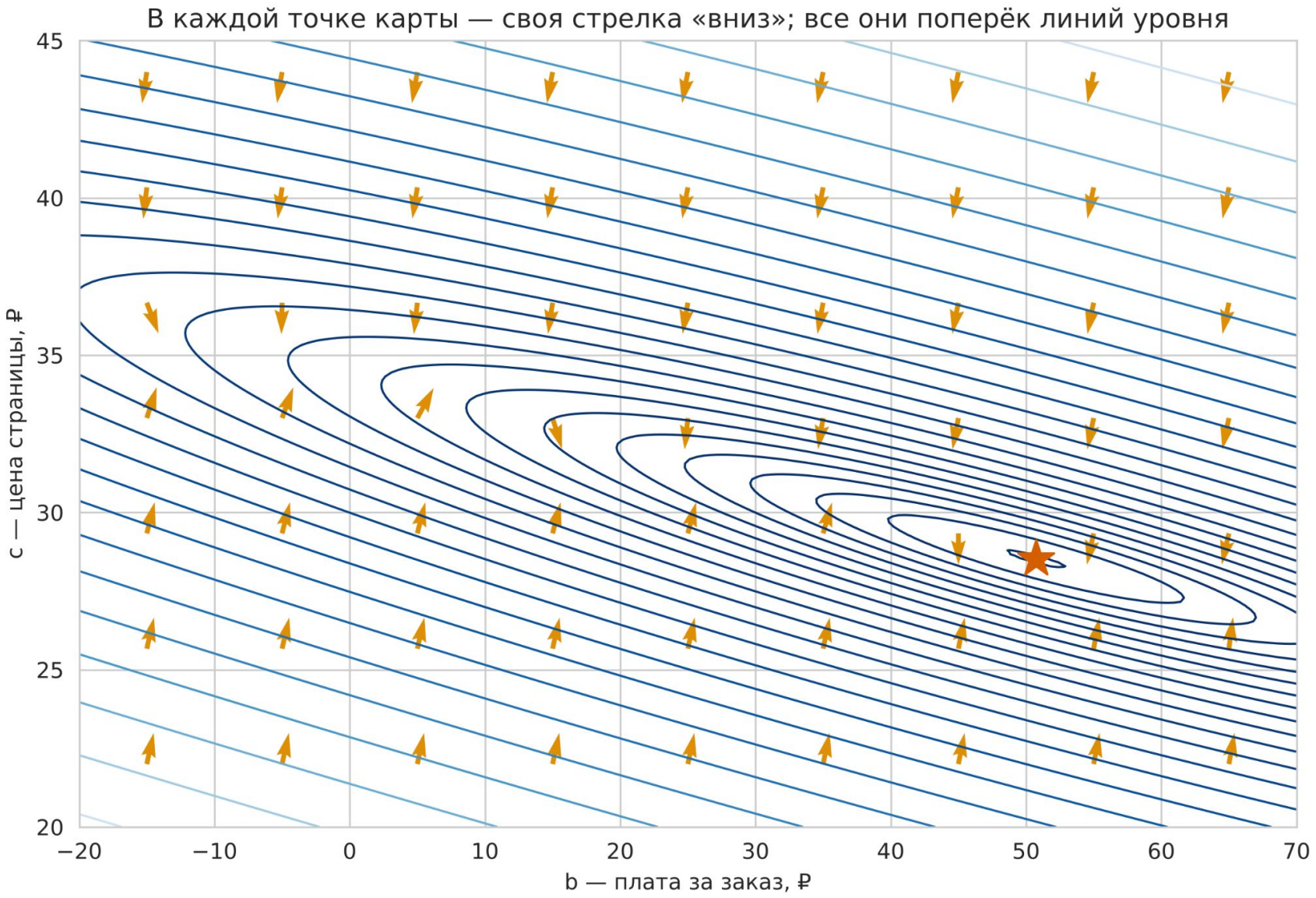
От одной переменной к нескольким



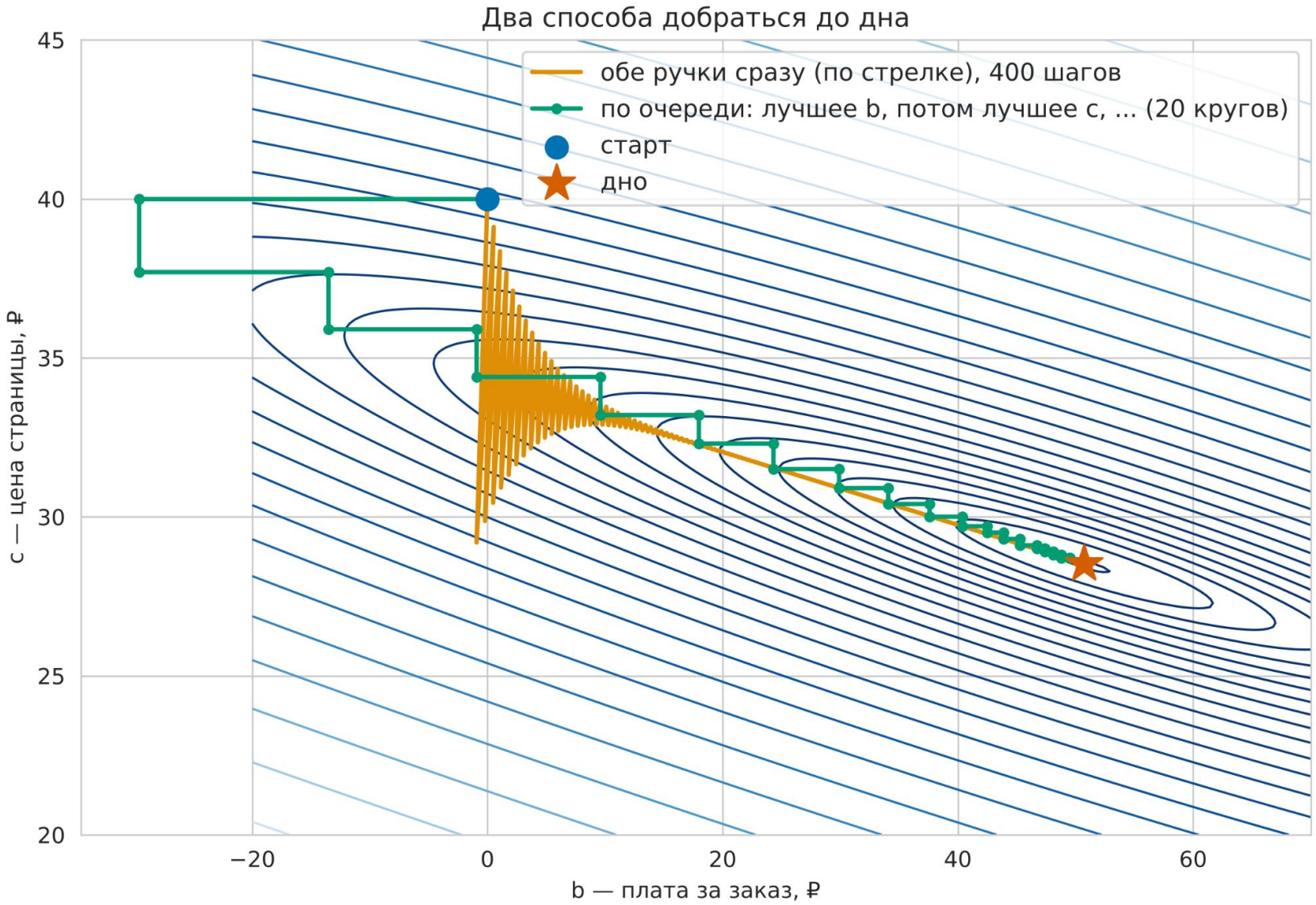


- Если несколько направлений, то производная всё так же определяется, но их теперь несколько
- **Градиент** – это вектор из частных производных
- Он показывает, в какую сторону функция быстрее всего возрастает

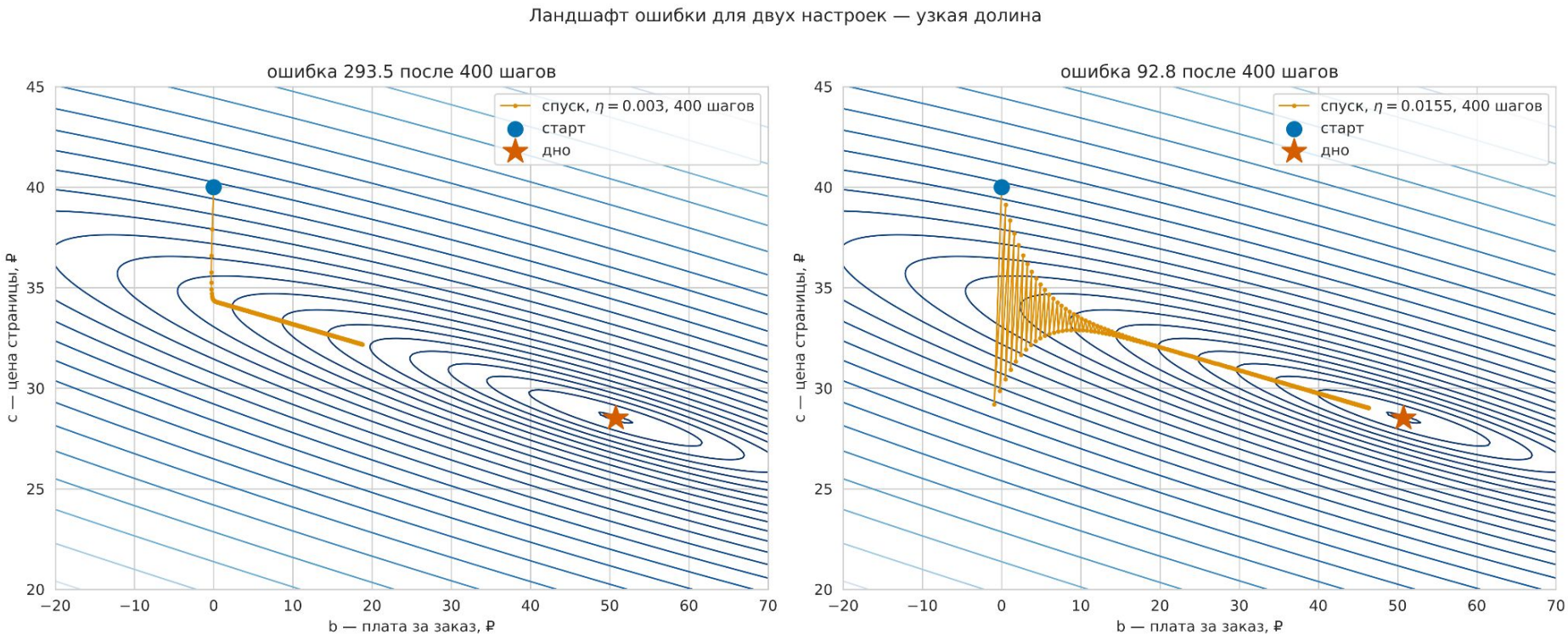
От одной переменной к нескольким



Можно было бы по очереди оптимизировать

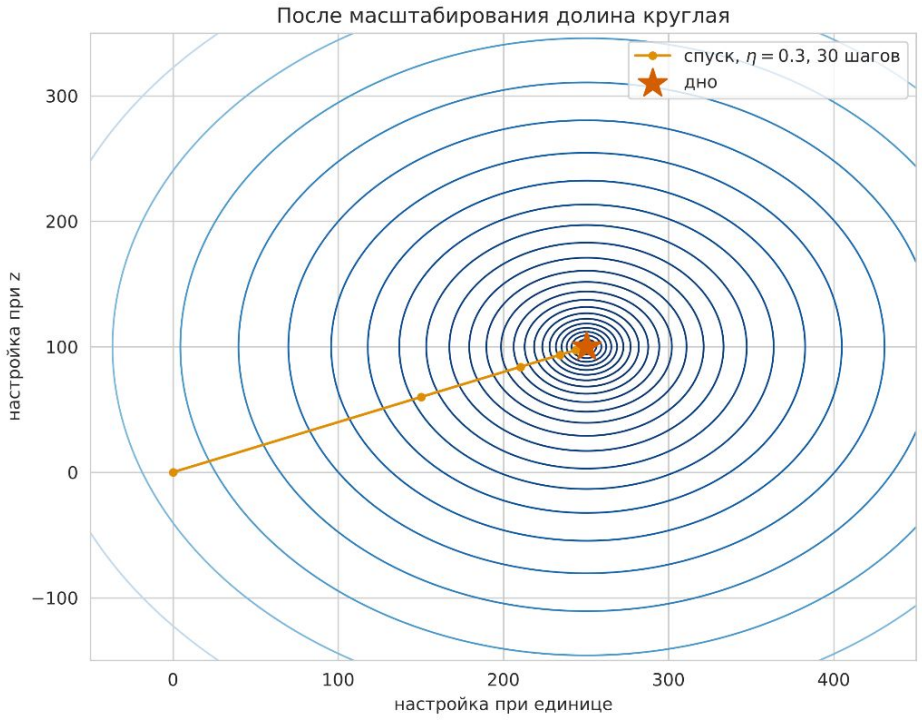


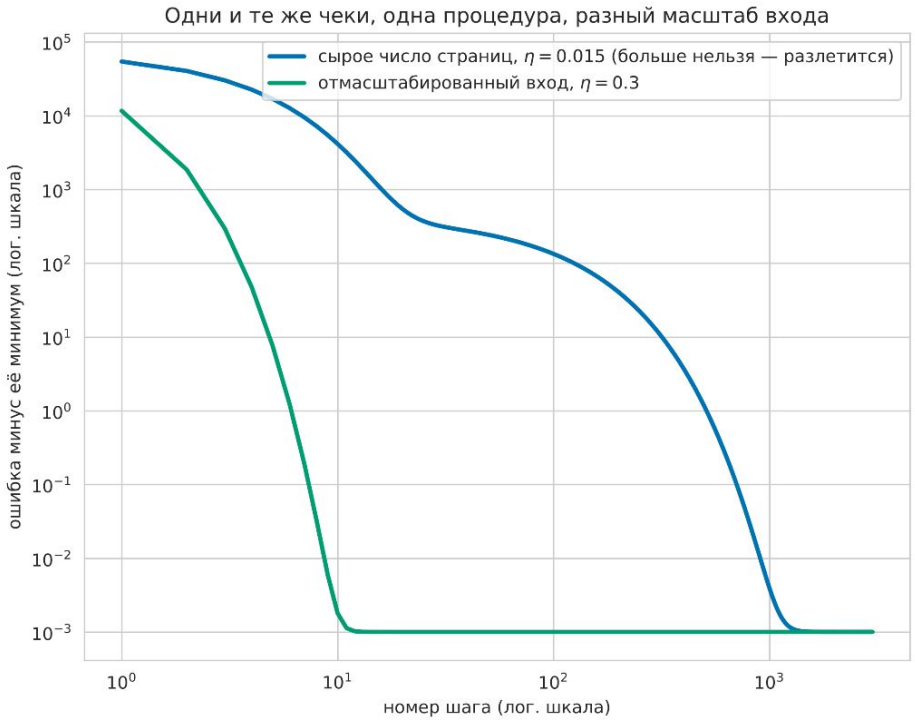
От одной переменной к нескольким



- Со скоростью обучения теперь всё ещё сложнее, потому что по разным направлениям ландшафт может быть совсем разный

От одной переменной к нескольким



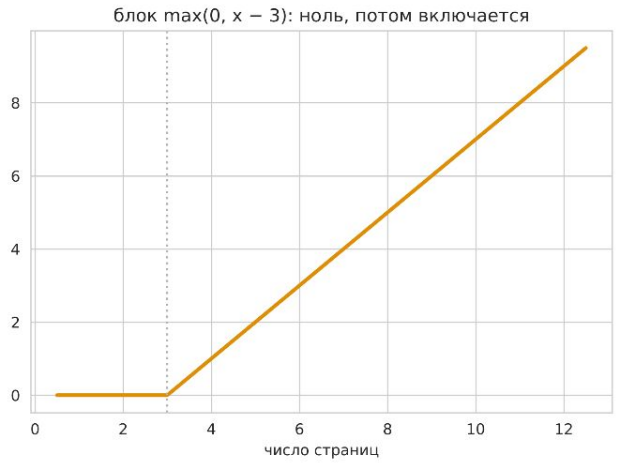


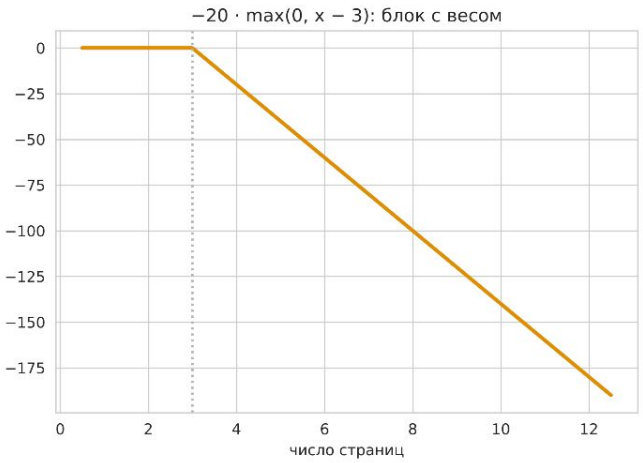
- Если бы можно было взять и перемасштабировать, жизнь стала бы куда проще; но это можно только на входе, а внутри сети жизнь всё равно будет сложна…

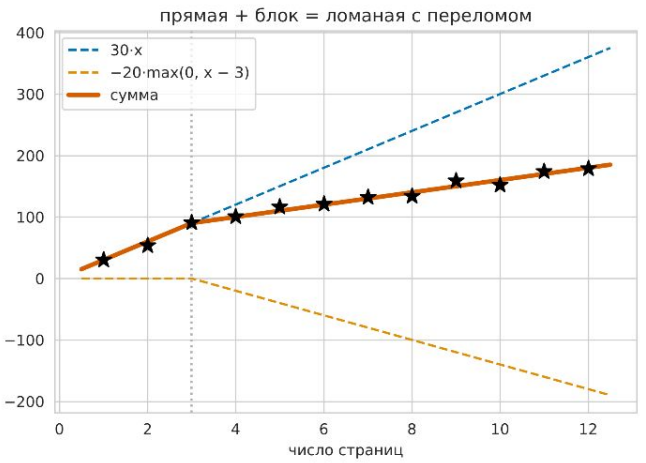
От наших примеров к нейросетям



Более сложное правило ценообразования



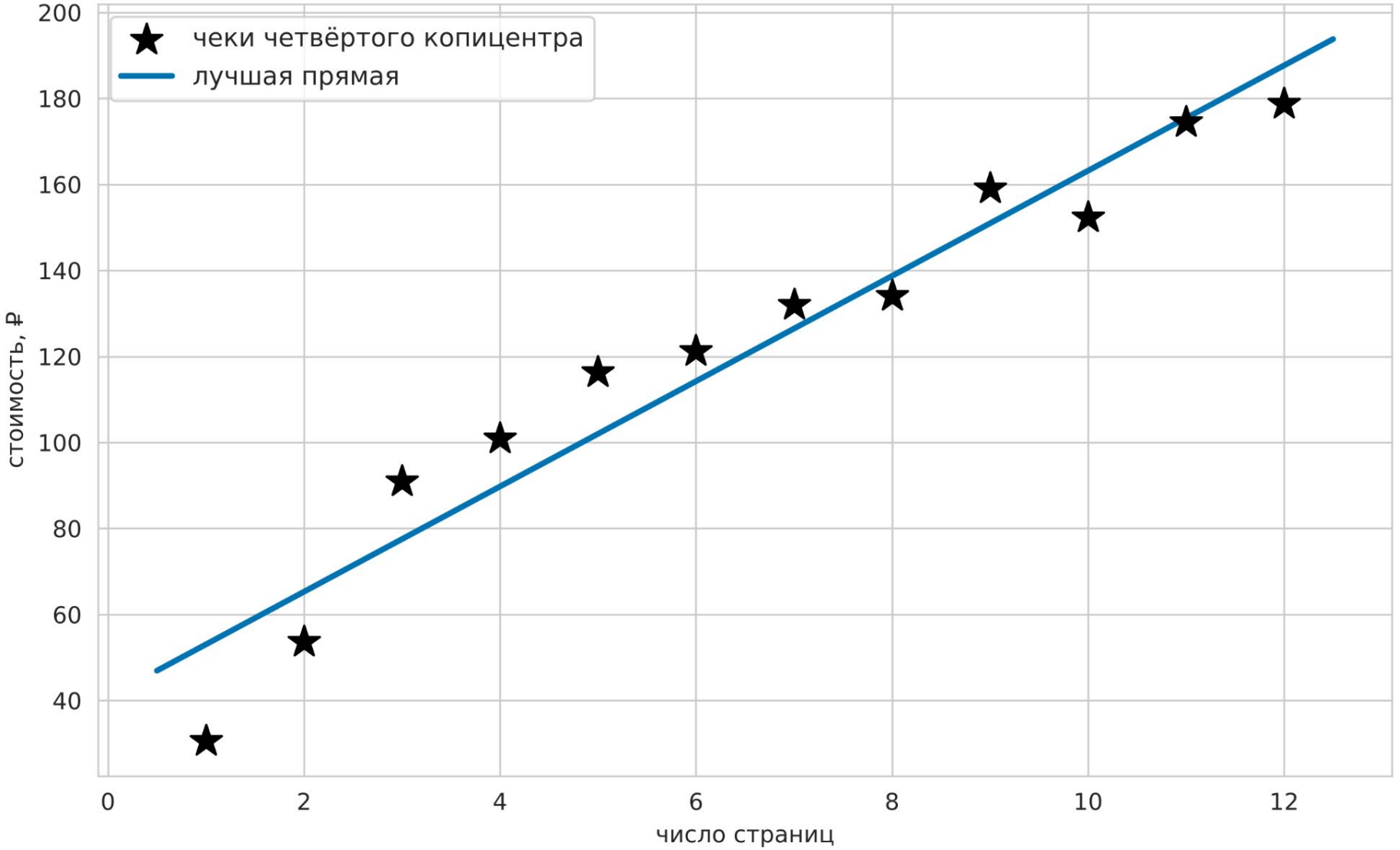




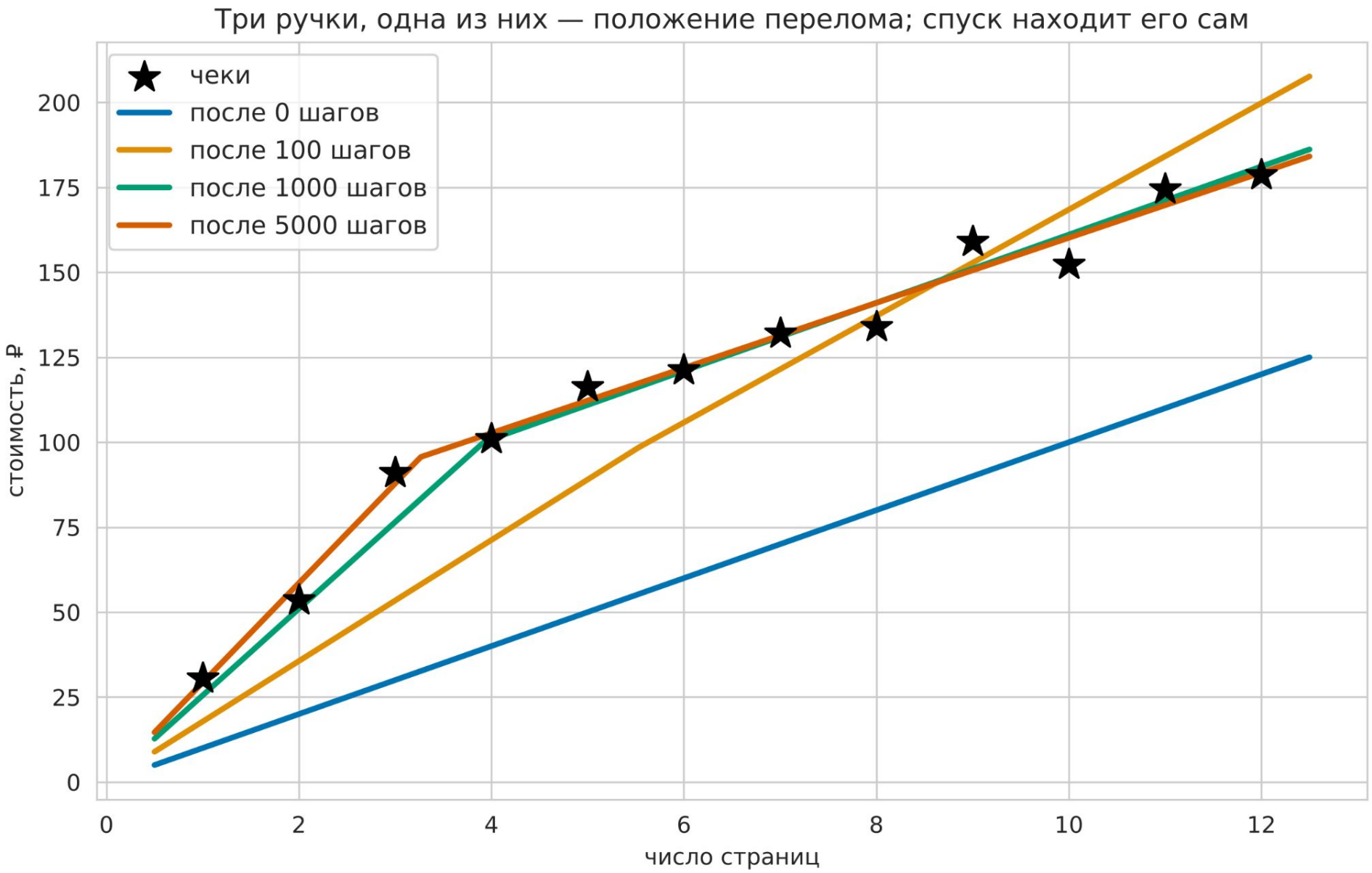
- Что если правило немного сложнее, чем было?
- Например, дают скидку за опт?..

Более сложное правило ценообразования

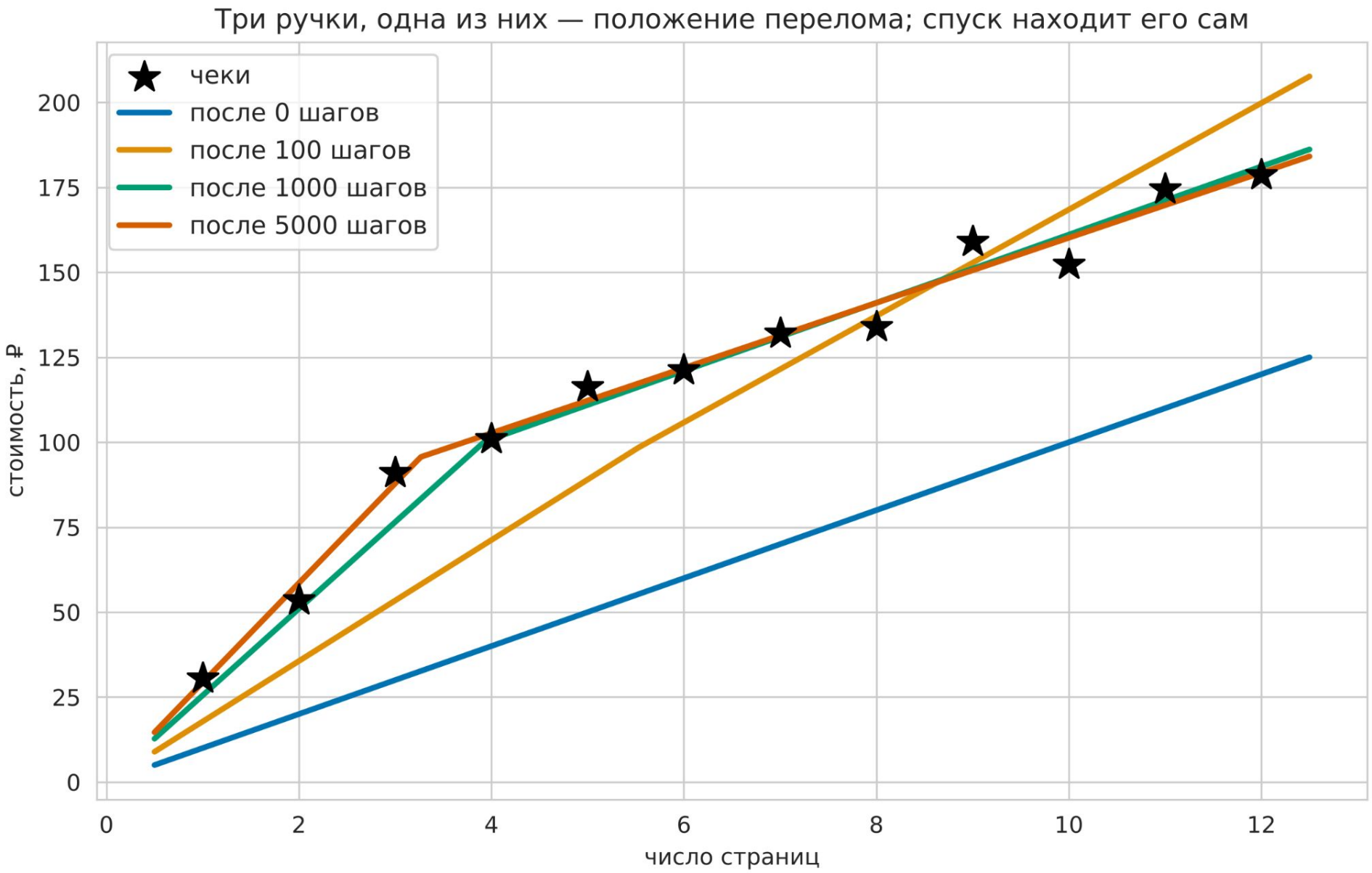
Прямая не справляется: до трёх страниц занижает, после — завышает, потом снова занижает



Более сложное правило ценообразования

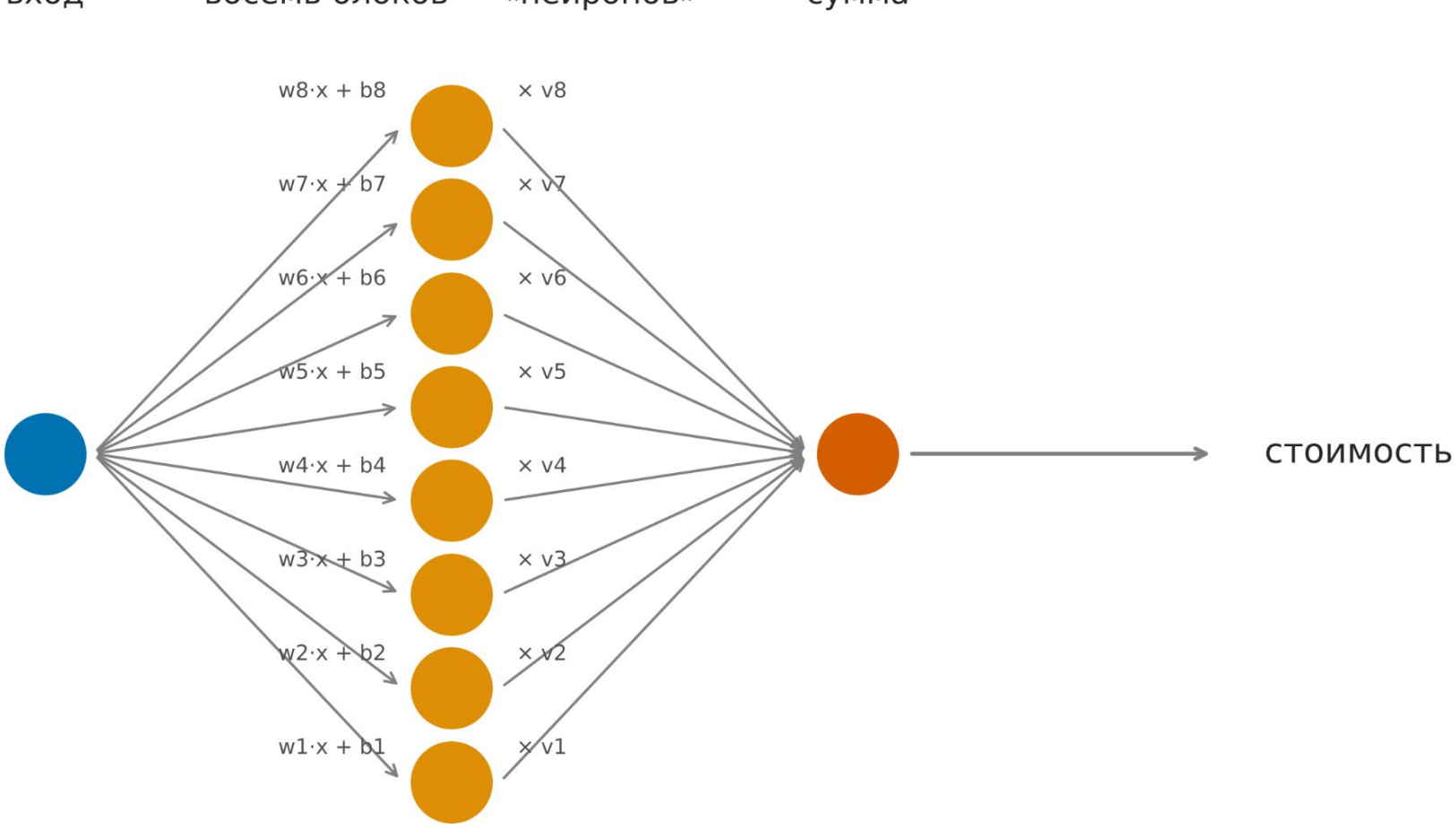


Более сложное правило ценообразования



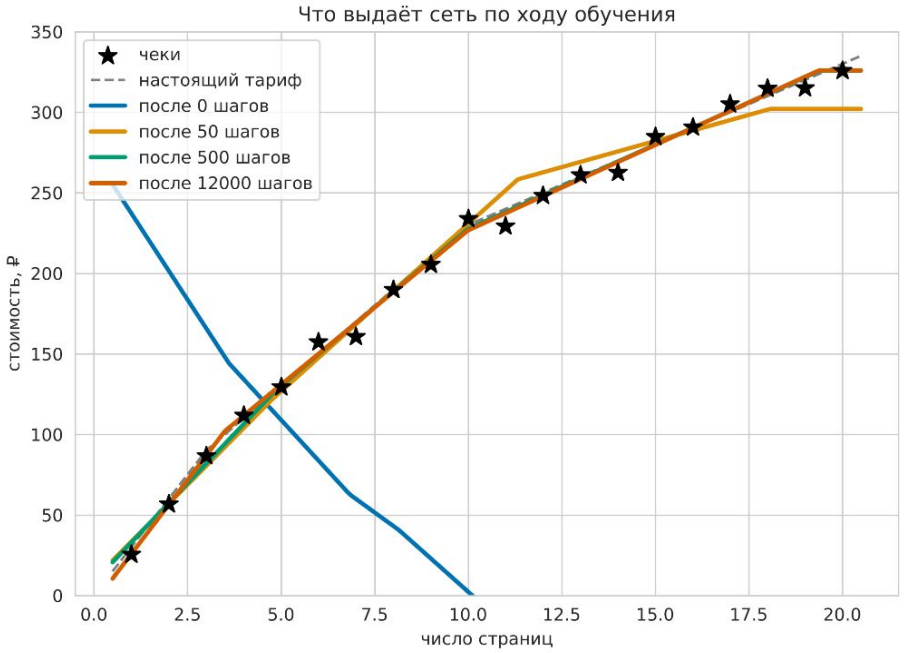
Более сложное правило ценообразования

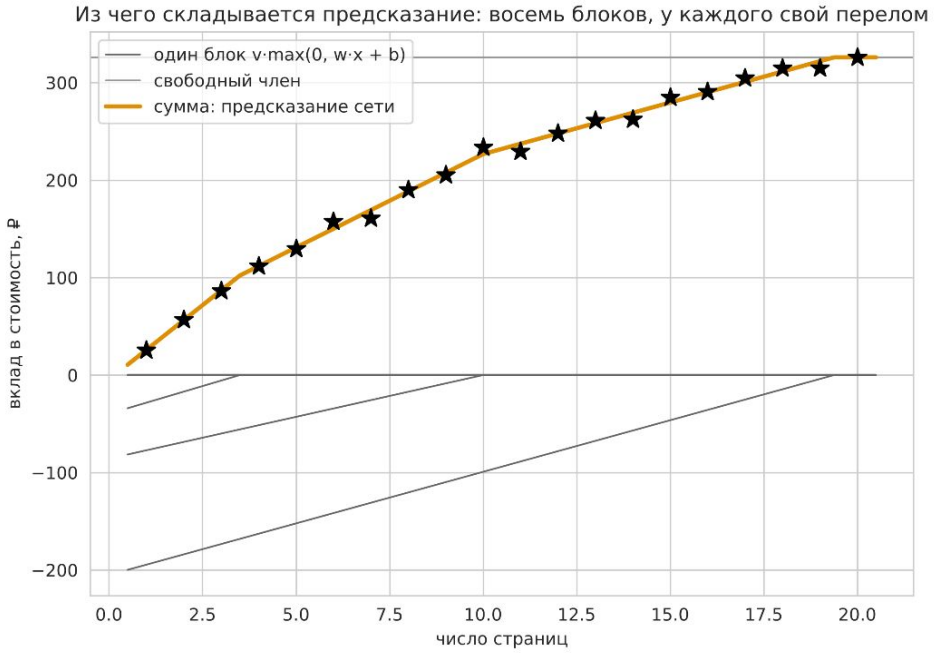
Это и есть нейросеть: один вход, восемь нейронов с ReLU, один линейный выход



- Если эту идею чуток обобщить, получится... как раз нейросеть с ReLU-активациями!

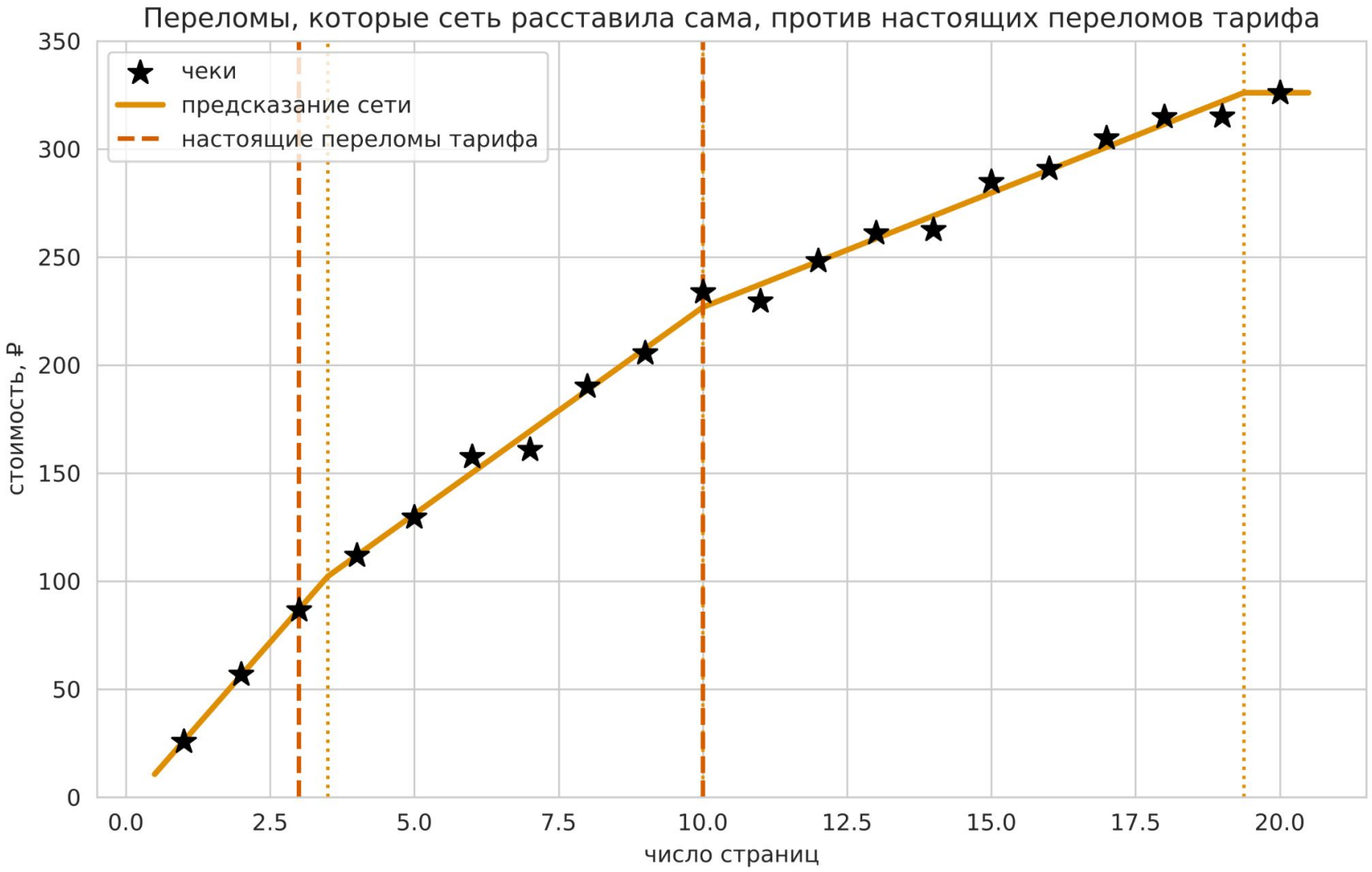
Более сложное правило ценообразования





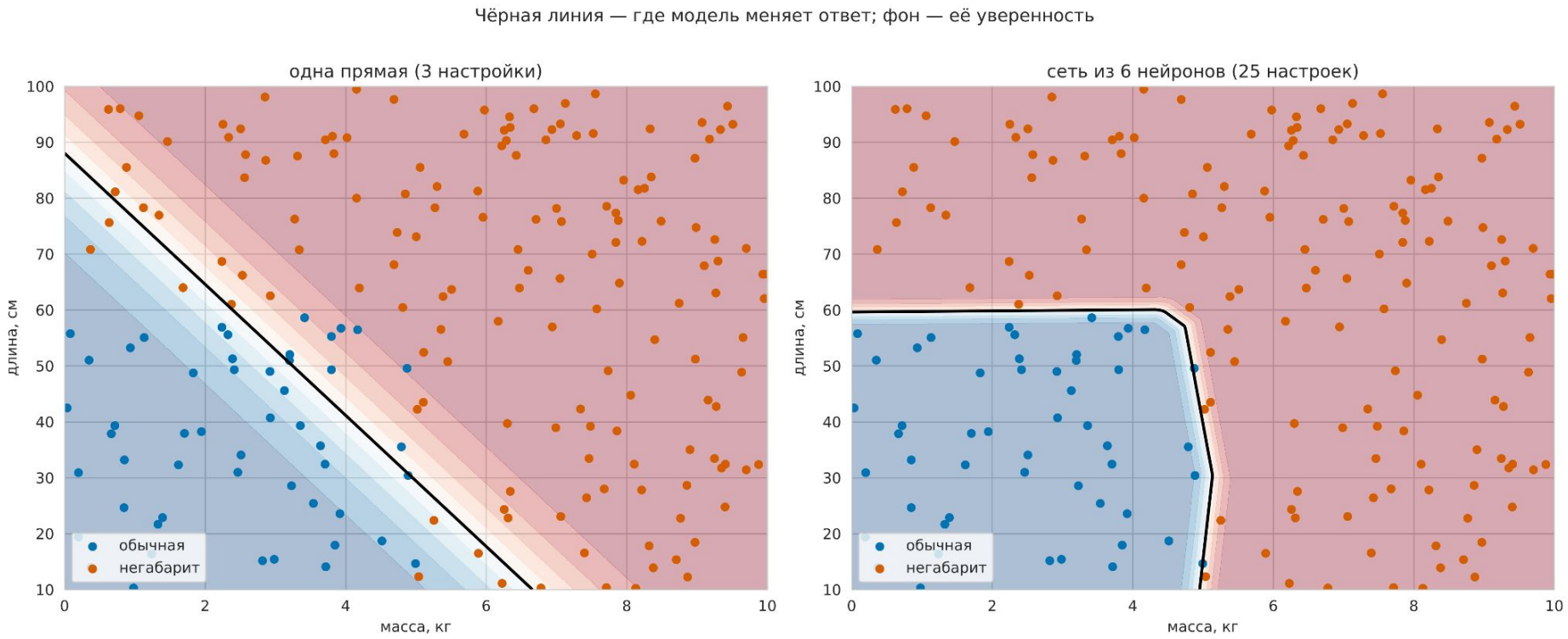
- И она будет обучаться и подбирать параметры для нужной ломаной

Более сложное правило ценообразования



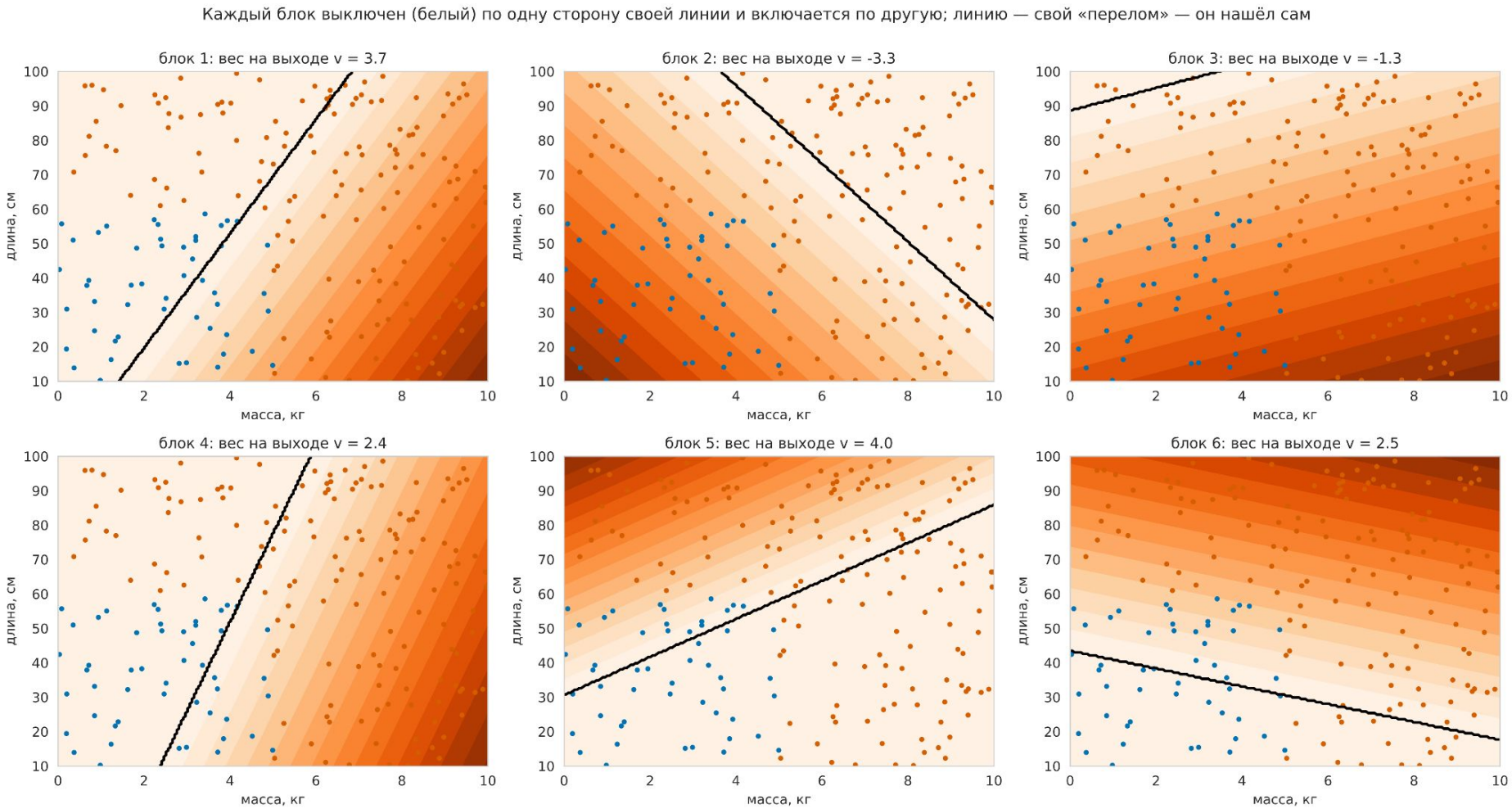
- Но здесь уже начнётся, конечно, оверфиттинг...

Классификация



- То же можно провернуть и 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

Южный федеральный университет

Спасибо за внимание!

@SINECOR



