

Линейная регрессия, L2 (Ridge) и L1 (Lasso)

Лазар В. И., Козлова Е. Р.

16 сентября 2025 г.

Постановка: линейная регрессия

Дано: выборка $\{(x_i, y_i)\}_{i=1}^n$, где $x_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$. **Модель:**

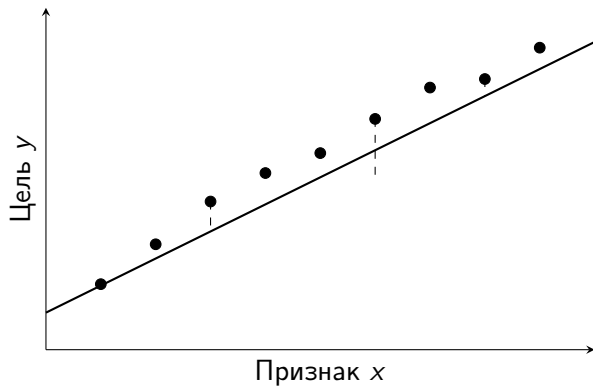
$$\hat{y} = \mathbf{w}^\top \mathbf{x} + b = \sum_{i=0}^d x_i w_i + b$$

Критерий (MSE):

$$\text{MSE}(\mathbf{w}, b) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Цель: найти \mathbf{w} , b , минимизирующие MSE и хорошо обобщающие на новых данных.

Интуиция MSE: «лучше всех приближает точки»



Остаток $r_i = y_i - \hat{y}_i$ — вертикальное отклонение точки от прямой. MSE усредняет квадраты остатков: большие промахи штрафуются сильнее. Линия «балансирует» точки, уменьшая суммарную квадратичную ошибку.

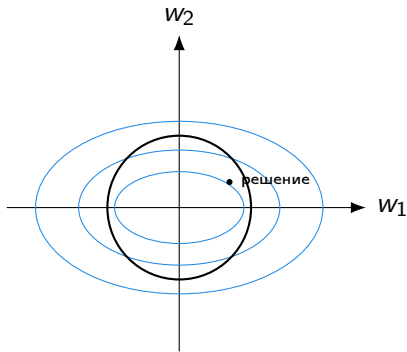
L2-регуляризация (Ridge): «сжимает» веса

Добавляем штраф на крупные веса:

$$J_{\text{ridge}}(w, b) = \text{MSE}(w, b) + \lambda \|w\|_2^2 = \text{MSE}(w, b) + \lambda \sum_{i=0}^d w_i^2.$$

Эффект: коэффициенты «сжимаются» к нулю, модель стабильнее к шуму и мультиколлинеарности. Обычно требуется **стандартизовать** признаки.

Ridge: геометрическая картина



Минимизируем MSE при ограничении $\|w\|_2 \leq R$. Касание эллипса уровня MSE окружности даёт решение: веса *умеренные* по длине.

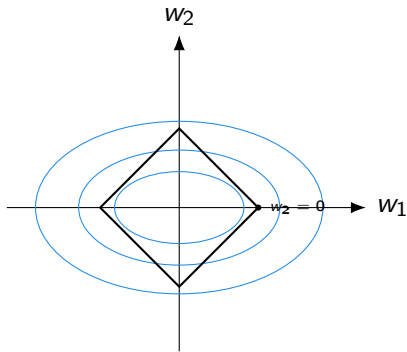
L1-регуляризация (Lasso): разреженность

Добавляем штраф на сумму модулей весов:

$$J_{\text{lasso}}(w, b) = \text{MSE}(w, b) + \lambda \|w\|_1 = \text{MSE} + \lambda \sum_i |w_i|.$$

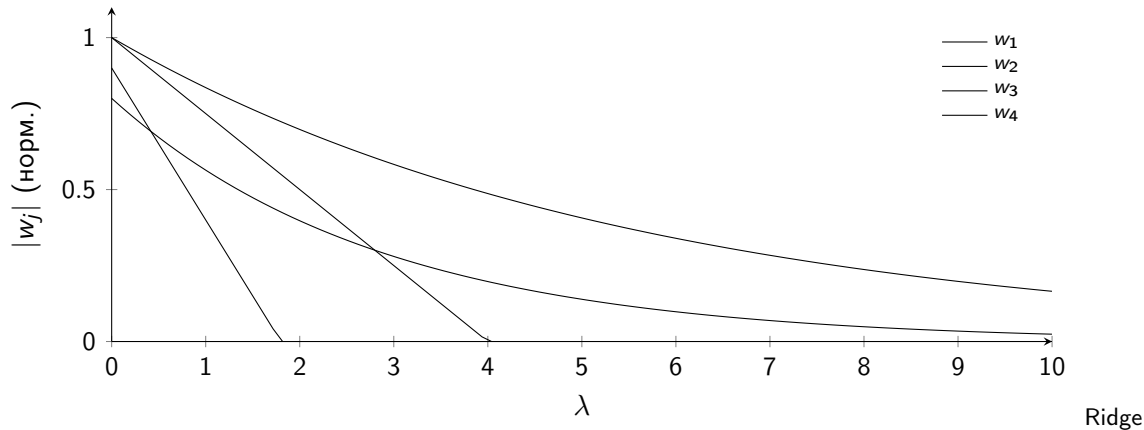
Эффект: часть коэффициентов становится ровно нулём (**разреженность**). Это помогает отбирать важные признаки.

Lasso: геометрическая картина



Из-за «углов» ромба решение часто попадает в вершину: некоторые w_j становятся ровно нулём — это и есть разреженность.

Как влияет λ на коэффициенты



плавно «сжимает» все веса; Lasso может занулять некоторые при умеренных λ .

Ridge vs Lasso: когда что лучше

- **Ridge** хорош, когда много слабых признаков и важно стабилизировать веса, нет жёсткой необходимости в отборе признаков.
- **Lasso** подходит, когда ожидаем, что среди признаков есть немногие важные: он делает модель разреженной (удобно для интерпретации).
- Можно комбинировать: **Elastic Net** $= \lambda_1 \|w\|_1 + \lambda_2 \|w\|_2^2$.

- Делите данные на train/validation/test; используйте кросс-валидацию для подбора λ .
- **Стандартизируйте признаки** перед регуляризацией: иначе штрафы несопоставимы.
- Проверяйте пере/недообучение по кривым валидации.
- Сохраняйте простоту: начинайте с Ridge, пробуйте Lasso, сравнивайте по валидации.