

ЛЕКЦИЯ 10. Регуляризация: Ridge-регрессия — как навести порядок в модели

Введение

Вы уже знаете, как строить регрессионную модель, включать в неё множество переменных и даже учитывать качественные признаки.

Вы также научились интерпретировать коэффициенты и проверять остатки.

Но вот парадокс:

Чем **больше** переменных вы добавляете, тем **лучше** становится R^2 ,

...но тем **хуже** модель работает на новых данных.

Это называется **переобучение (overfitting)** — когда модель **не выявляет закономерности**, а просто **запоминает особенности выборки**.

Признаки переобучения:

- R^2 близок к 1
- Остатки при проверке модели на новых данных — огромные
- Коэффициенты скачут, меняют знак
- Модель "боится" новых данных

Что делать?

- Нужно **ввести ограничение** на поведение коэффициентов.
- Нужно, чтобы модель **не слишком полагалась на каждый признак**.
- Это и есть **регуляризация**.

Что такое регуляризация?

Это метод, который **штрафует** модель за слишком большие коэффициенты.

Зачем?

- Чтобы модель была **более устойчива**,
- Чтобы она **не переоценивала случайные переменные**,
- Чтобы она **предсказывала**, а не просто “запоминала” данные.

Ridge-регрессия (ℓ_2 -регуляризация)

Это вид линейной регрессии с регуляризацией, в котором добавляется штраф за **величину коэффициентов**.

 Формула:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \lambda * (b_1^2 + b_2^2 + \dots + b_n^2)$$

Где:

- λ (лямбда) — коэффициент регуляризации:

→ чем он больше, тем сильнее “сжимаются” коэффициенты

- $(b_1^2 + b_2^2 + \dots + b_n^2)$ — штраф за сложность модели

Как это работает?

Без регуляризации	С Ridge-регуляризацией
Модель подстраивается под каждый шум	Модель сглаживает шум
Коэффициенты могут быть огромные	Коэффициенты компактные
Высокий риск переобучения	Устойчивость выше

Пример

Вы строите модель, объясняющую успеваемость на основе:

количества сна

уровня стресса

мотивации

количества кружков

удовлетворённости жизнью

и ещё 10 переменных

! Все они хоть как-то коррелируют между собой.

→ Модель становится хрупкой, переменные “перетягивают канат”.

Ridge-регрессия:

удерживает коэффициенты под контролем

снижает переобучение

делает модель пригодной для новых данных

Что делает λ (лямбда)?

Значение λ	Что происходит с моделью
$\lambda = 0$	Обычная линейная регрессия
$\lambda \rightarrow \text{большое}$	Коэффициенты $\rightarrow 0$, модель “уплощается”
$\lambda = \infty$	Все коэффициенты становятся ≈ 0

Ваша задача — найти оптимальное λ , чтобы модель была:

- не слишком гибкой,
- но и не слишком жёсткой

Обычно его подбирают с помощью кросс-валидации (об этом позже).

Как сделать Ridge в Excel/Sheets?

Классическая LINEST не поддерживает Ridge.

Но вы можете:

1. Ввести обычную модель
2. Добавить вручную **штраф** в отдельную ячейку:

$$\lambda * (b_1^2 + b_2^2 + \dots + b_n^2)$$

3. Визуально сравнить обычную модель и “приглушённую” модель

Для полной реализации лучше использовать Python/Google Colab — если тебе интересно, могу потом показать, как это подключить для студентов.

Когда применять Ridge?

Ситуация	Подходит Ridge?
Много переменных (10+)	✓ Да
Высокая корреляция между переменными	✓ Да
Нестабильные коэффициенты	✓ Да
Модель слишком простая	✗ Нет

Что можно сделать вручную (в вашем курсе)

- Постройте обычную модель
- Проверьте: не слишком ли большие коэффициенты?
- Добавьте “визуально” Ridge-штраф в анализ
- Объясните: какие переменные могли быть “переоценены”

Использование ИИ

Инструмент	Что делает
ChatGPT	Объясняет логику λ , когда применять Ridge
Excel Copilot	Подсказывает, где модель перегружена
Notion AI	Помогает описать логику регуляризации в текстовом блоке

🚫 Запрещено:

- Применять Ridge без объяснения, зачем
- Не сравнивать результат “до и после”
- Писать “мы использовали Ridge”, если этого не было
- Подставлять λ наугад без интерпретации

Вывод

Ridge-регрессия — это способ **воспитать модель**.

Она не позволяет ей быть слишком агрессивной, не даёт ей “увлечься” шумом.

Это шаг к тому, чтобы ваша модель была **устойчива, строгая, и настоящая аналитическая работа — начиналась**.