



Методы регрессионного анализа

Юлия Пономарева
Data Scientist

Проверка связи



Отправьте «+», если меня видно и слышно

Если у вас нет звука или изображения:

- перезагрузите страницу
- попробуйте зайти заново
- откройте трансляцию в другом браузере (используйте Google Chrome или Microsoft Edge)
- с осторожностью используйте VPN, при подключении через VPN видеопотоки могут тормозить

Цели занятия

1. Вспомним основные метрики для задач регрессии
2. Рассмотрим обучение линейной регрессии
3. Обсудим полиномиальную регрессию
4. Изучим регуляризацию

План занятия



1. Метрики регрессии
2. Линейная регрессия
3. Полиномиальная регрессия
4. Регуляризация
5. Итоги занятия

Метрики регрессии

Метрики регрессии

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}|$$

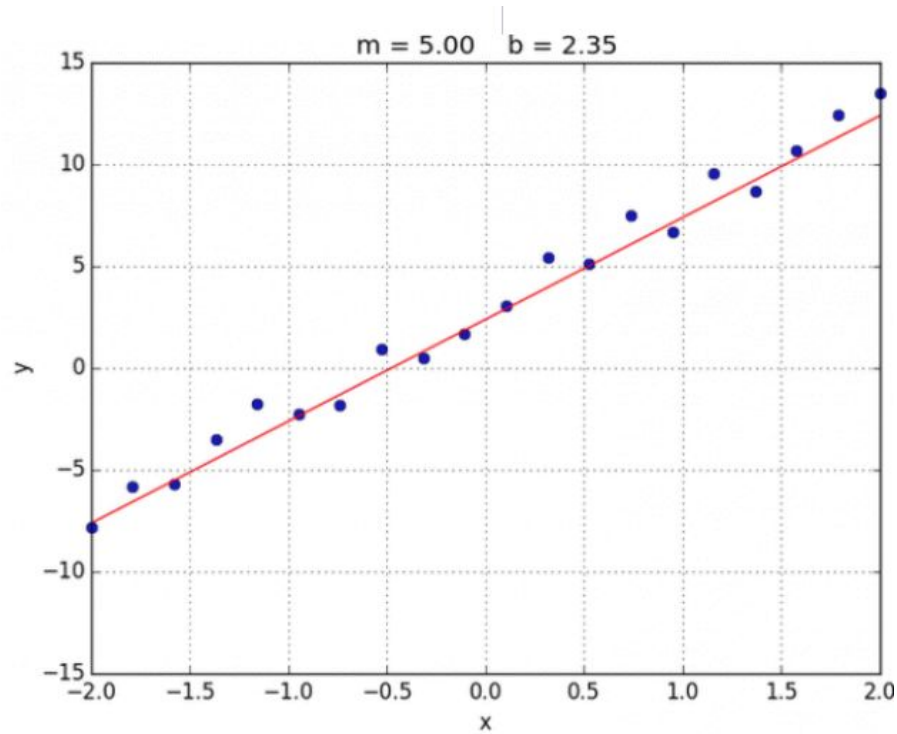
$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2}$$

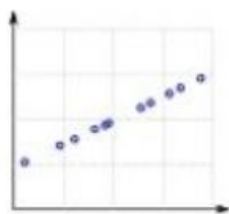
Линейная регрессия

Линейная регрессия



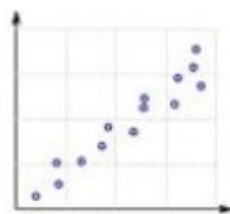
Требование к данным

Высокая скоррелированность признаков



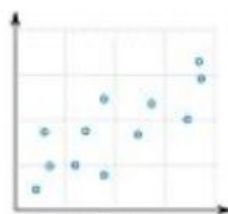
1

Максимальная
положительная
корреляция



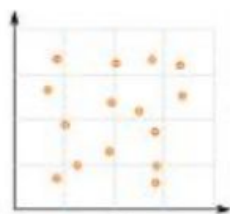
0.8

Высокая
положительная
корреляция



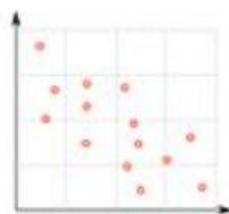
0.3

Низкая
положительная
корреляция



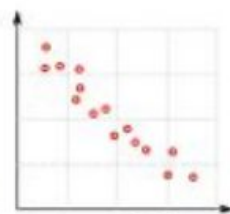
0

Отсутствие
корреляции



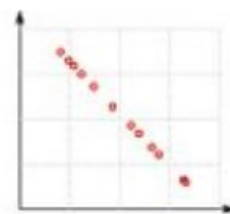
-0.3

Низкая
отрицательная
корреляция



-0.8

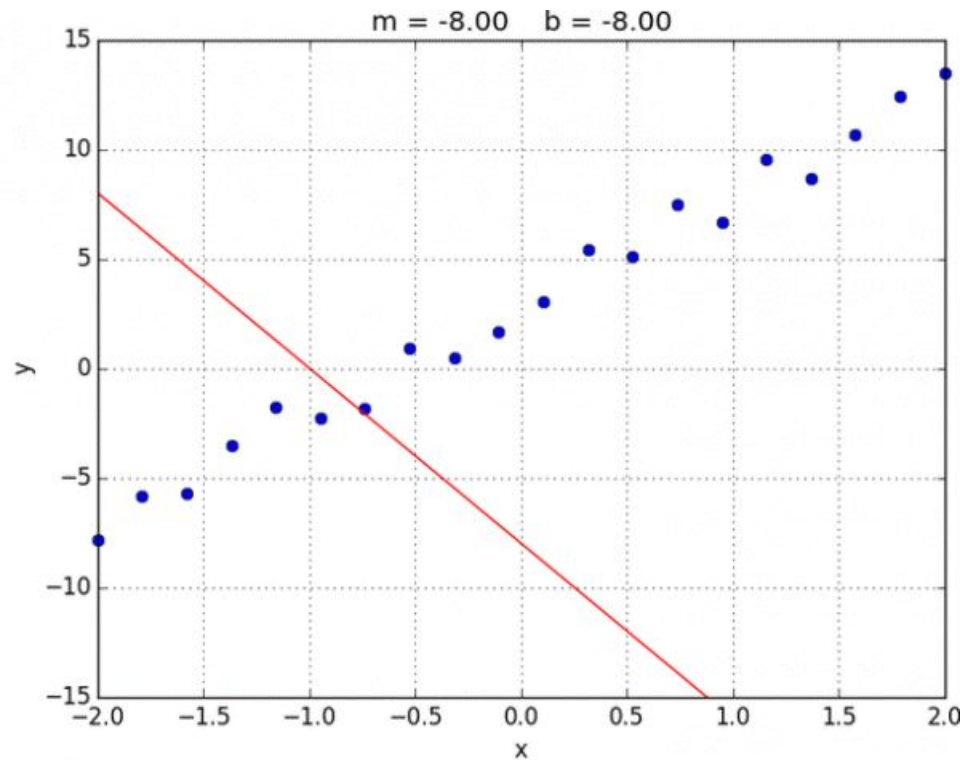
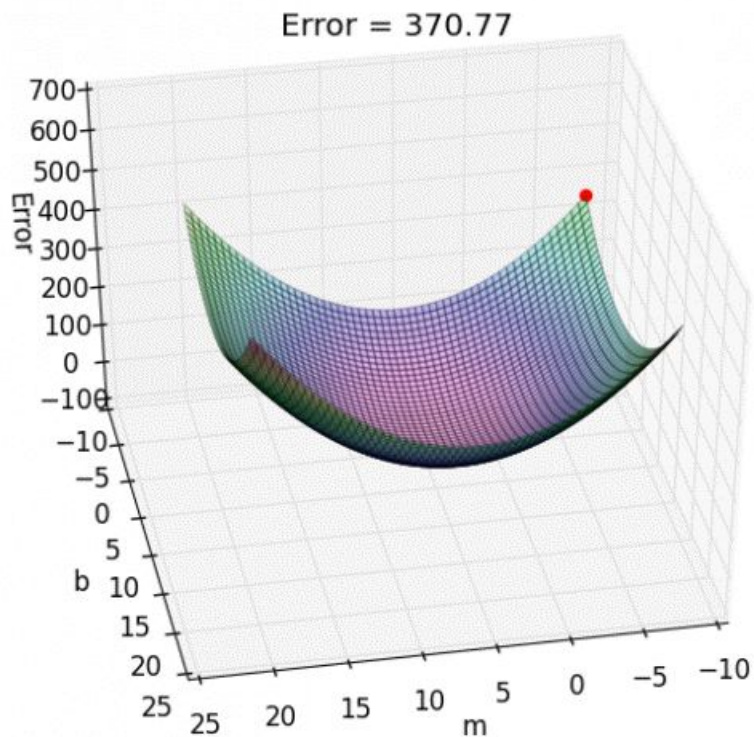
Высокая
отрицательная
корреляция



-1

Максимальная
отрицательная
корреляция

Линейная регрессия (обучение)



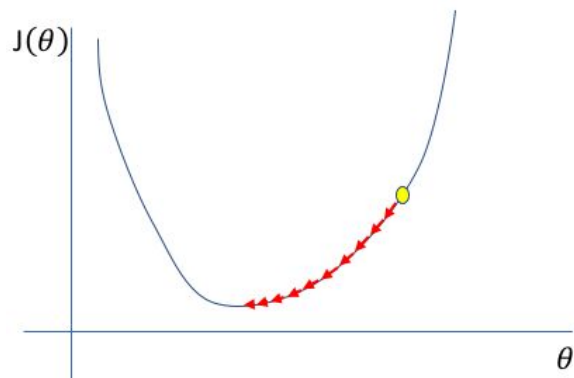
Линейная регрессия (обучение)

1. Выбираем начальное приближение w
2. Цикл по $k = 1, 2, 3 \dots$:

- $s_k = -\nabla Q(w^{k-1}, X)$
- $w^k = w^{k-1} + \eta_k s_k$
- Если $\|w^k - w^{k-1}\| < \epsilon$, то завершить

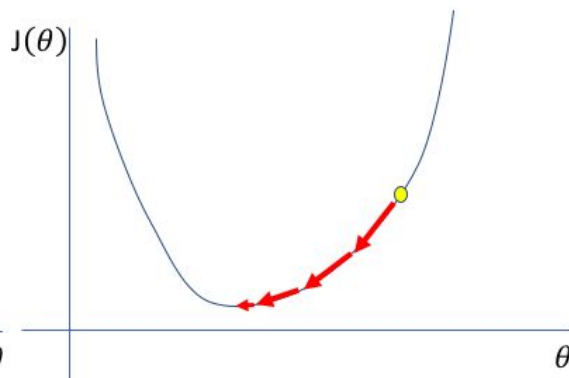
Скорость градиентного спуска

Too low



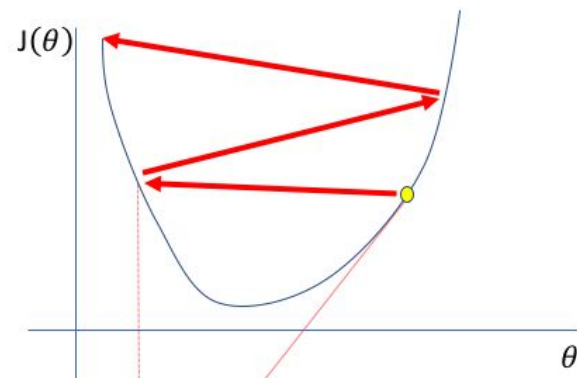
A small learning rate requires many updates before reaching the minimum point

Just right



The optimal learning rate swiftly reaches the minimum point

Too high



Too large of a learning rate causes drastic updates which lead to divergent behaviors

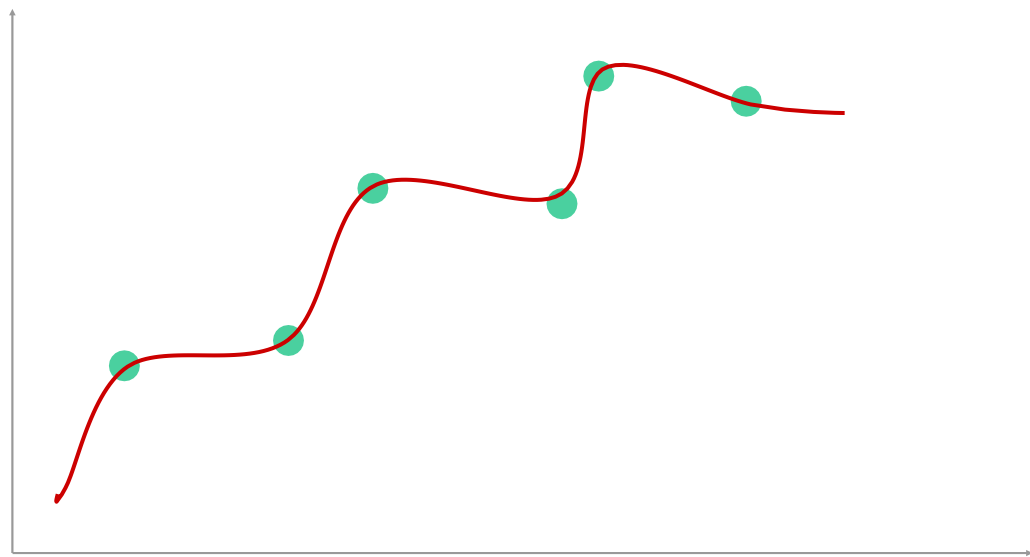
Практика

(Линейная регрессия)

(Градиентный спуск)

Полиномиальная регрессия

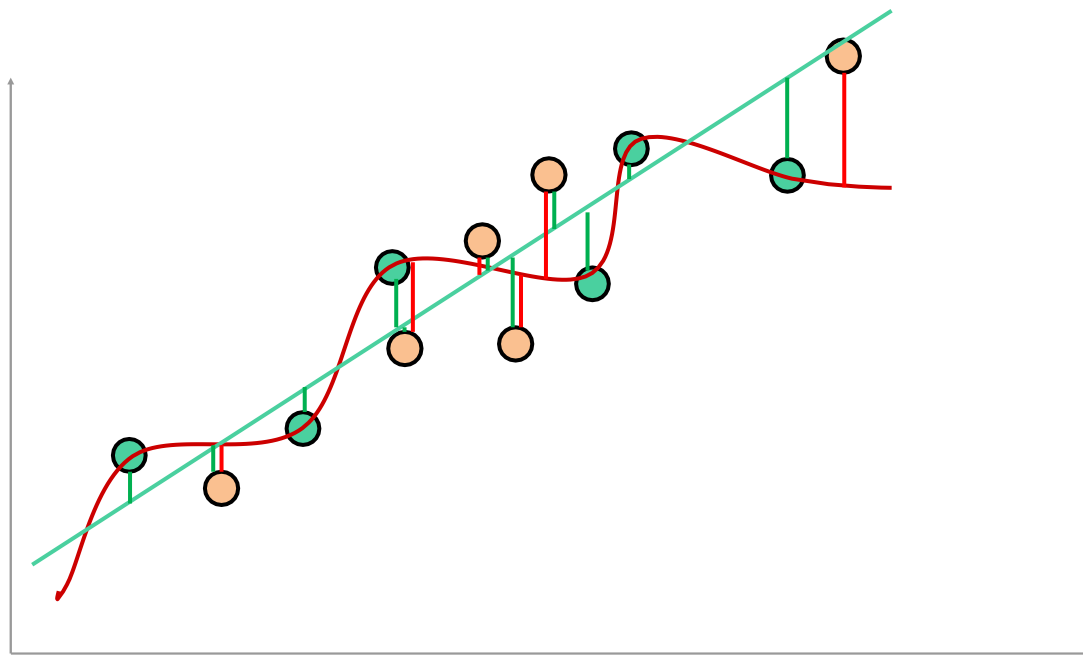
Полиномиальная регрессия



$$y = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + a_5x^5$$

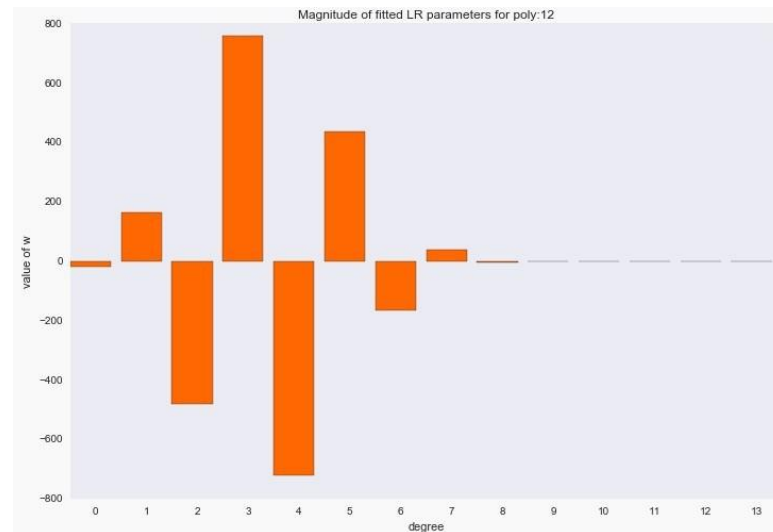
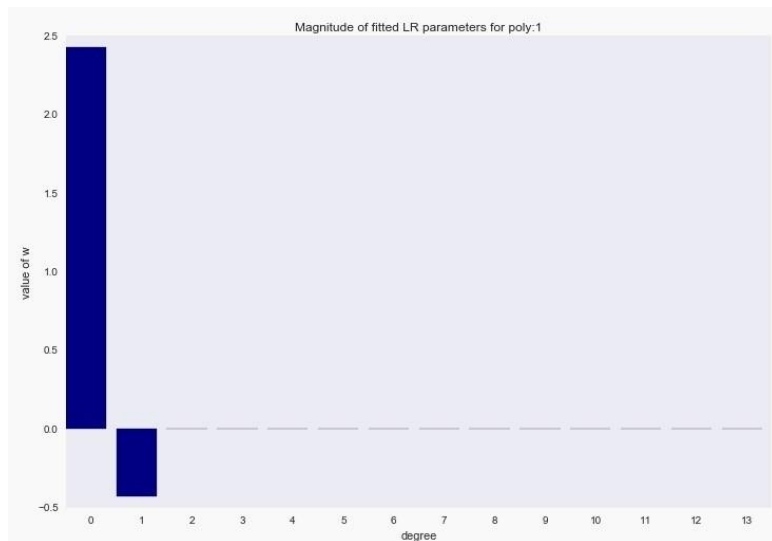
Полиномиальная регрессия

На тестовых данных получим большую ошибку:



Полиномиальная регрессия

При увеличении степени полинома коэффициенты быстро растут



Регуляризация

L1-регуляризация (lasso, регуляризация через манхэттенское расстояние)

$$L_1 = \sum_{i=1} |y_i - \widehat{y}_i| + \lambda \sum_{j=1} |w_j|$$

L2-регуляризация (ridge, регуляризация Тихонова)

$$L_2 = \sum_{i=1} (y_i - \widehat{y}_i)^2 + \lambda \sum_{j=1} w_j^2$$

Чем больше λ , тем меньшая сложность модели будет получаться в процессе обучения:

- если увеличивать его, в какой-то момент оптимальным для модели окажется зануление всех весов
- при слишком низких его значениях появляется вероятность чрезмерного усложнения модели и переобучения

λ подбирают по метрикам

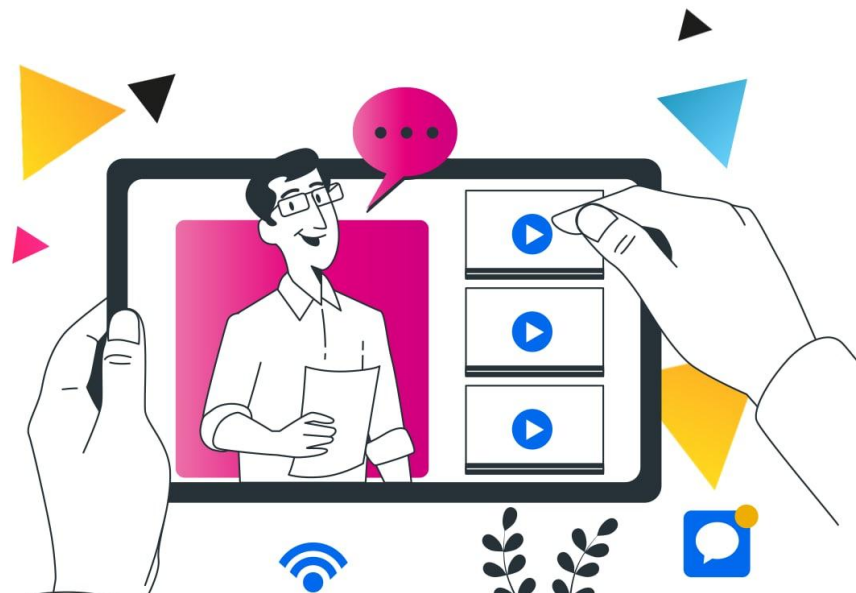
$$\begin{cases} Q(a, X) \rightarrow \min \\ ||a||^2 \leq C \end{cases}$$

Практика

(Полиномиальная регрессия)

(Регуляризация)

Ваши вопросы?



Итоги занятия

Итоги занятия



1. Вспомнили основные метрики для задач регрессии
2. Рассмотрели обучение линейной регрессии
3. Обсудили полиномиальную регрессию
4. Изучили регуляризацию

1. Реализация линейной регрессии
<https://youtu.be/KJA9A1q9I7E>
2. Метрики MSE, MAE, R2 <https://youtu.be/vh2smjQyhp8>
3. Регуляризация в линейной модели -
https://www.youtube.com/watch?v=L_o8v5A23XA
4. Регуляризация -
<https://propoprogs.ru/ml/ml-l2-regularizator-matematicheskoe-obosnovanie-i-primer-raboty>
5. Регуляризация -
<https://www.youtube.com/watch?v=RvcP3a727Fg>
6. Регуляризация - <https://youtu.be/Q81RR3yKn30>

Пожалуйста, оставьте
свой отзыв о семинаре

[ссылка](#)



До встречи!

