Линейная регрессия

регуляризация

Переобучение

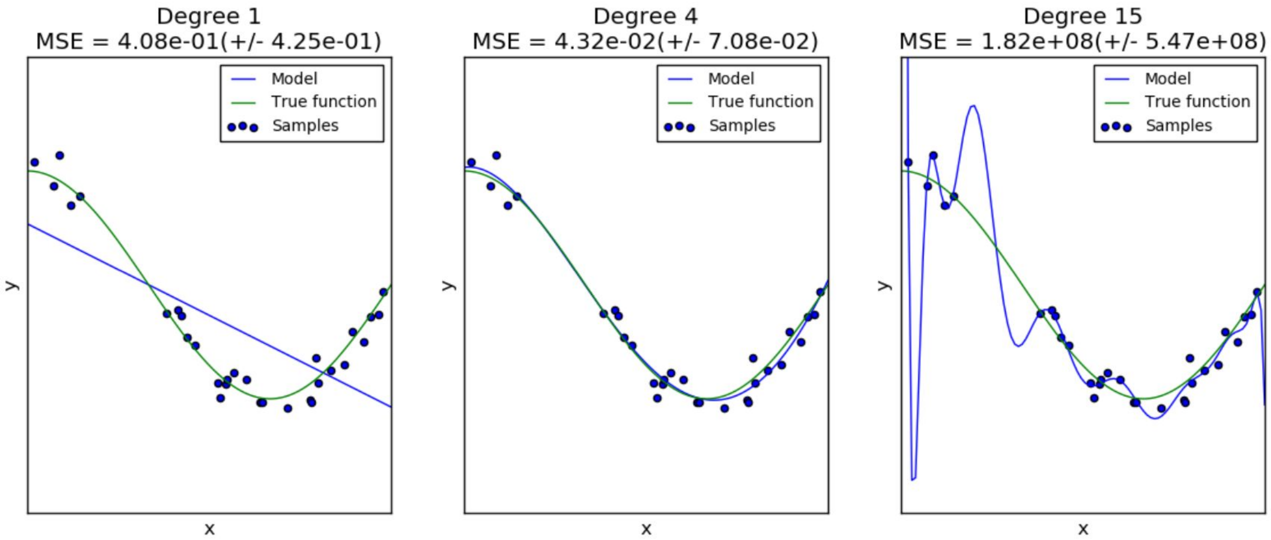


Рис. 1. Регрессионные кривые для признаковых наборов различной сложности.

https://github.com/esokolov/ml-course-hse/blob/master/2023-fall/lecture-notes/lecture03-linregr.pdf

Отложенная выборка

- По обучающей выборке нельзя отследить переобучение
- Разделяем данные на две части train/validation
- Обучаем модель на обучающей части, на валидационной проверяем качество

Проблема:

Результат существенно зависит от конкретного разбиения данных на обучение и контроль

Cross Validation

	Test	Train	Train	Train	Train
	Train	Test	Train	Train	Train
	Train	Train	Test	Train	Train
	Train	Train	Train	Test	Train
	Train	Train	Train	Train	Test
Errors	120.55	122.11	125.91	123.41	122.81

Mean Error 122.96

Обучение линейной регрессии

$$\frac{1}{l}\sum_{i=0}^{l}(\langle w, x_i\rangle \; - \; y_i)^2 \rightarrow \underset{w}{min}$$

$$\frac{1}{l}||Xw \; - \; y||^2 \; \rightarrow \; \underset{w}{min}$$

$$w \; = \; \left(X^T X\right)^{-1} X^T y$$

Обучение линейной регрессии

Аналитическое решение

$$w = \left(X^T X\right)^{-1} X^T y$$

Проблемы

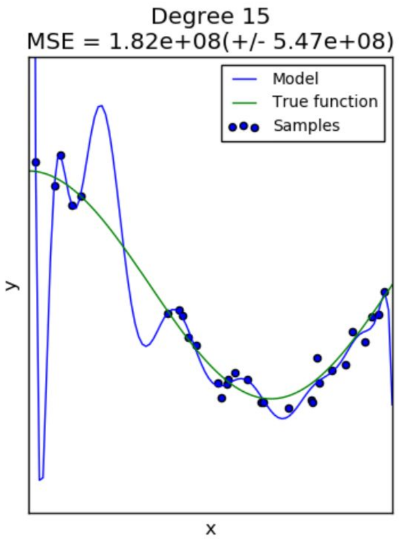
- Матрица $X^T X$ может быть вырождена или плохо обусловлена
- Обращение матрицы сложная операция

Регуляризация

Пусть в выборке есть л.з. признаки

$$\exists\, v : \forall\, x\, \langle v,\, x\rangle \;=\; 0$$

$$\langle w \,+\, \alpha v,\, x\rangle \;=\; \langle w\,,\, x\rangle \,+\, \alpha\langle v,\, x\rangle \;=\; \langle w,\, x\rangle$$



Большие веса w симптом переобучения

т. к модель чувствительна к крайне малым изменениям признака

Регуляризация

$$\frac{1}{l}||Xw - y||^2 \rightarrow \underset{w}{min}$$

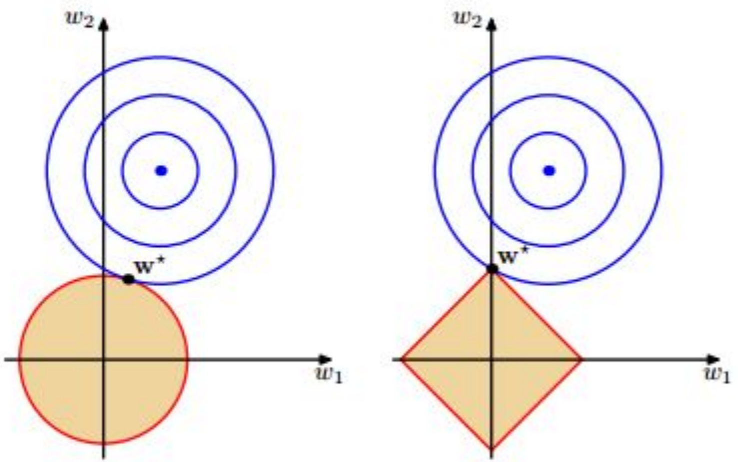
$$L_1 \qquad \frac{1}{l}||Xw - y||^2 + \alpha||w||_1 \rightarrow \underset{w}{min}$$

$$L_2 \qquad \frac{1}{l}||Xw - y||^2 + \alpha||w||_2^2 \rightarrow \underset{w}{min}$$

Регуляризация

На практике оказывается, что L_1 зануляет часть параметров модели,

а L_2 нет



Регуляризация

Нужно ли включать вес w_0 в регуляризатор?

$$w_0 + \frac{1}{l}\sum_{i=1}^{l}(\langle w, x_i\rangle - y_i)^2 \to \min_{w}$$