

ВЕБИНАРНЫЙ ФОРМАТ
Библиотеки Python для Data Science: продолжение

Урок 4.
Способы улучшения полученной модели.

Светлана Медведева

План урока

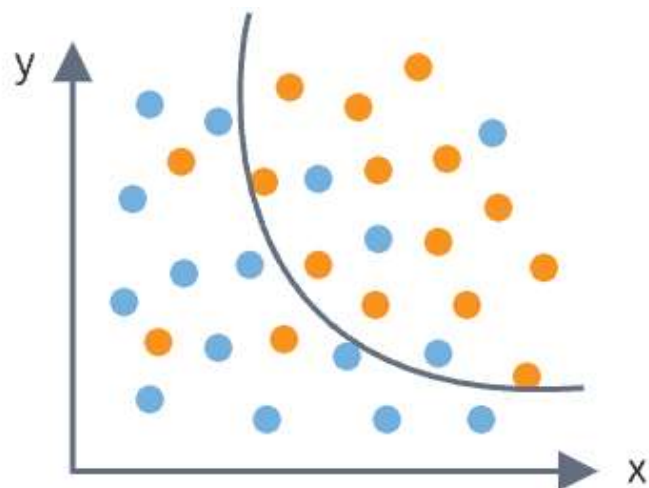
- Переобучение и способы борьбы с ним
- Регуляризация
- Практическая часть

Регуляризация

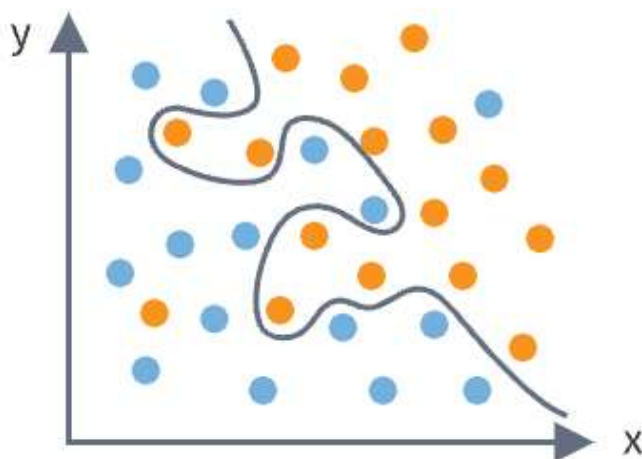
- Обнаружить переобучение:

- отложенная выборка
- кросс-валидация
- меры сложности модели

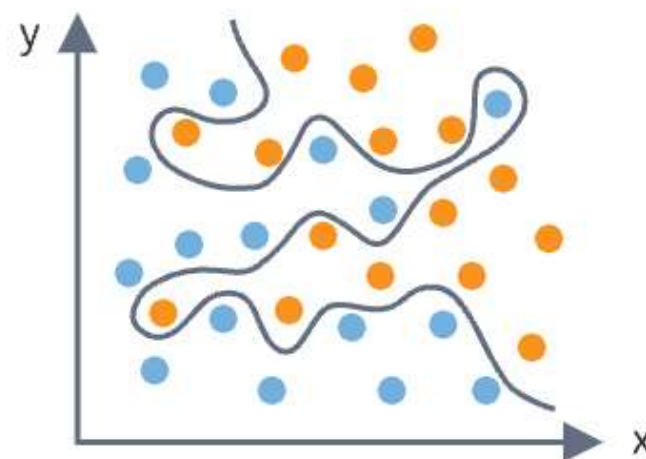
- Взять больше данных
- Выбрать более простую модель
- Регуляризация



Недообучение



Оптимум



Переобучение

Регуляризация

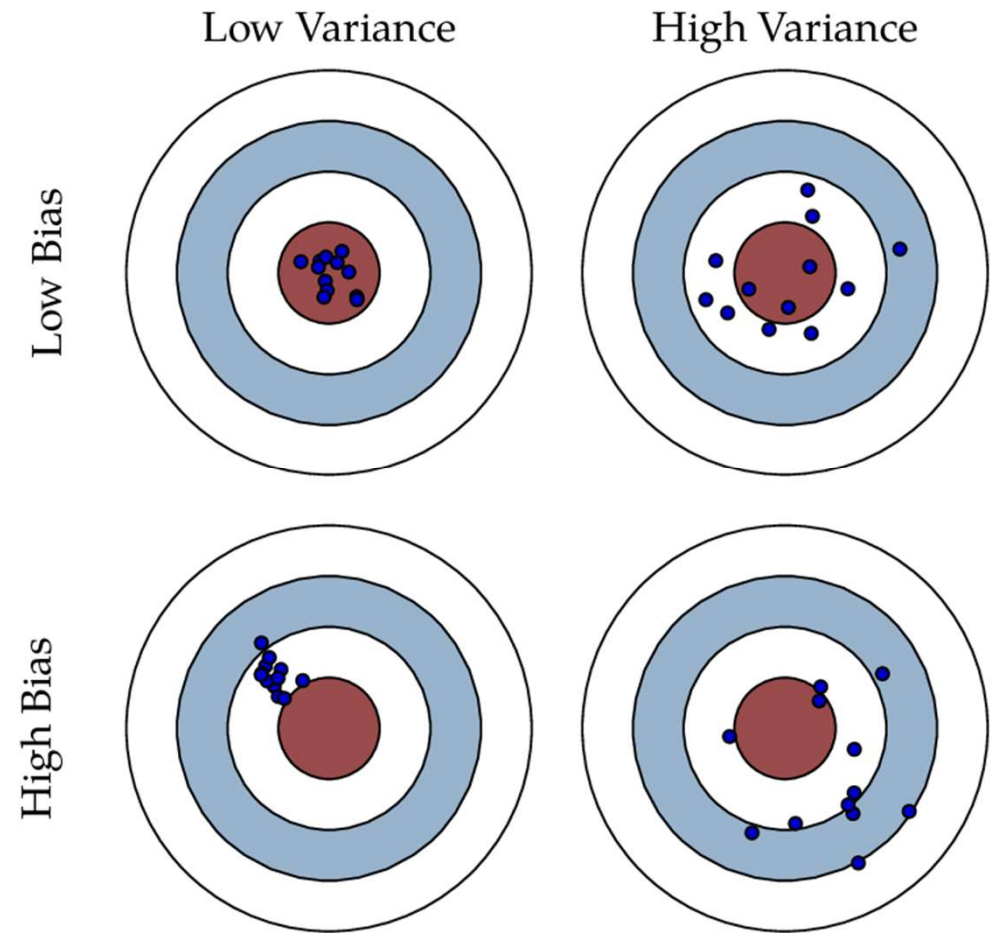
Регуляризация - метод борьбы с переобучением, с помощью добавления некоторого штрафа за сложность модели.

L1-регуляризация:

$$L_1 = \sum_{i=1} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1} |w_j|$$

L2-регуляризация:

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



Итоги урока

- Переобучение и способы борьбы с ним
- Регуляризация
- Практическая часть

Практическая часть

Спасибо за внимание!