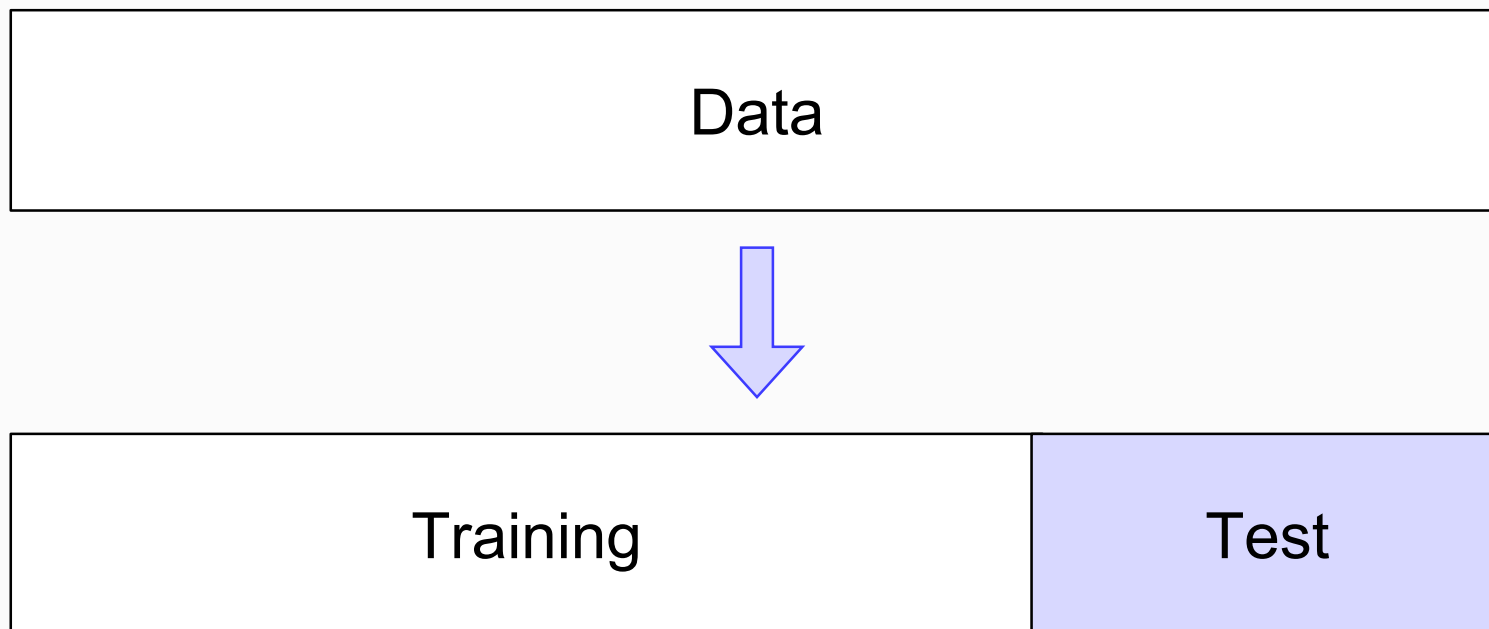


Кросс-валидация

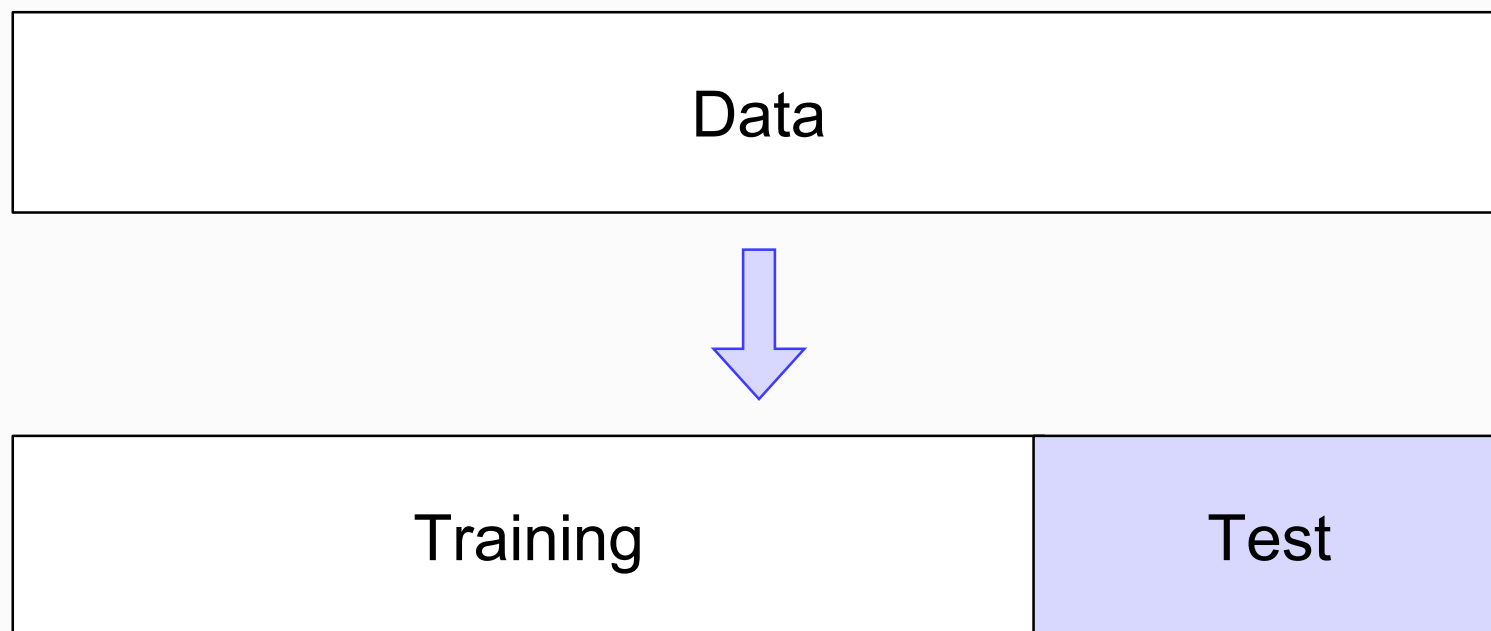
Елена Кантонистова

Skillbox

Отложенная выборка

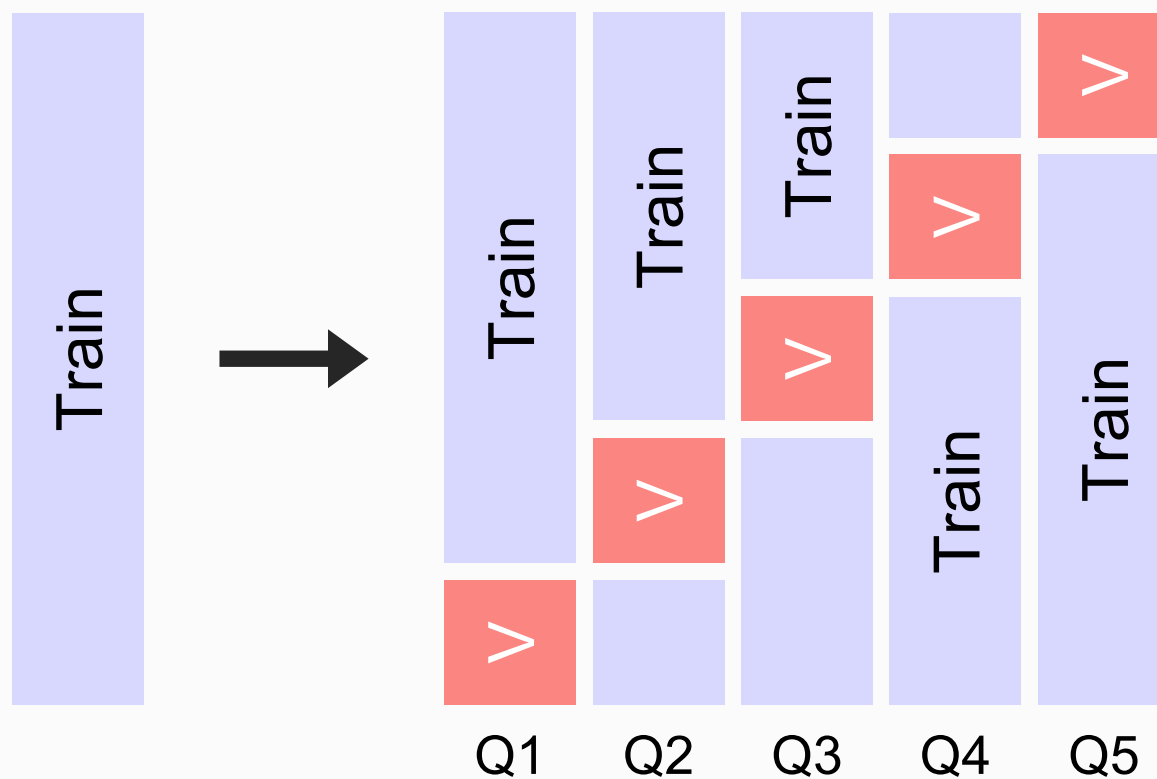


Отложенная выборка



Недостаток: качество модели сильно зависит от тех данных, на которых она обучалась. Значение метрик качества также сильно зависит от тестовых данных.

K-fold кросс-валидация



$$Q = \frac{1}{k} \sum_{i=1}^k Q_i$$

Где k — число блоков (фолдов).

K-fold кросс-валидация

- Среднее значение полученных метрик — это оценка качества модели, не привязанная к конкретному разбиению на train и test
- Стандартное отклонение: большое стандартное отклонение может свидетельствовать о переобучении

Пример

- Решается задача классификации и измеряется значение метрики roc-auc
- Результаты по 5-fold кросс-валидации получились следующие

Значения метрик на фолдах: [0.7, 0.5, 0.6, 0.85, 0.95]:

- $\text{mean } Q = (0.7+0.5+0.6+0.85+0.95)/5 = 0.72$
- $\text{std } Q = \sqrt{(0.7-0.72)^2+(0.5-0.72)^2+(0.6-0.72)^2+(0.85-0.72)^2+(0.95-0.72)^2} = 0.08$

Число блоков

В большинстве задач рекомендуется использовать от 3 до 5 блоков.

Если число блоков равно числу объектов выборки, то это leave-one-out кросс-валидация.



Вы узнали

Что такое k-fold кросс-валидация и почему качество модели лучше оценивать с помощью неё, а не при помощи отложенной выборки.