

Машинное обучение

Лекция 3. Обучение и оценка качества

Цель

Хотим, чтобы модель работала хорошо

Что такое хорошо?

- Функция потерь
- Метрики качества

На что можем влиять?

- Данные
- Признаки
- Модель
- Метод обучения

Данные

Зафиксируем признаки:

- Достаточно ли данных?
- Как их собирать?
- Как и кто будет размечать?

Задача

Собрали данные, определились с признаками, функцией потерь, метриками качества и выбрали модель. Надо её обучить и понять, хорошо ли модель будет работать на практике

Способы оценки качества

- Black-box методы
 - Online
 - Offline
- Glass-box методы
 - VC-оценки
 - PAC Bayes bounds
 - ...

Online

- Наблюдение
 - Эксперимент
-
- + В условиях эксплуатации
 - + В положительные результаты эксперимента обычно верят
 - + Легко хвастаться результатом
-
- Не всегда “боевые условия” доступны
 - Вряд ли цель эксплуатации
 - Можно навредить пользователям
 - Люди не любят быть объектом экспериментирования
 - Качество эксперимента может быть сильно хуже продакшена
 - Количество одновременных экспериментов ограничено

Offline

- + Нельзя навредить пользователям
- + Обычно можно проводить сильно больше экспериментов
- Обычно нужны данные (примеры)
- Сложно «хвастаться» результатом

Offline на данных

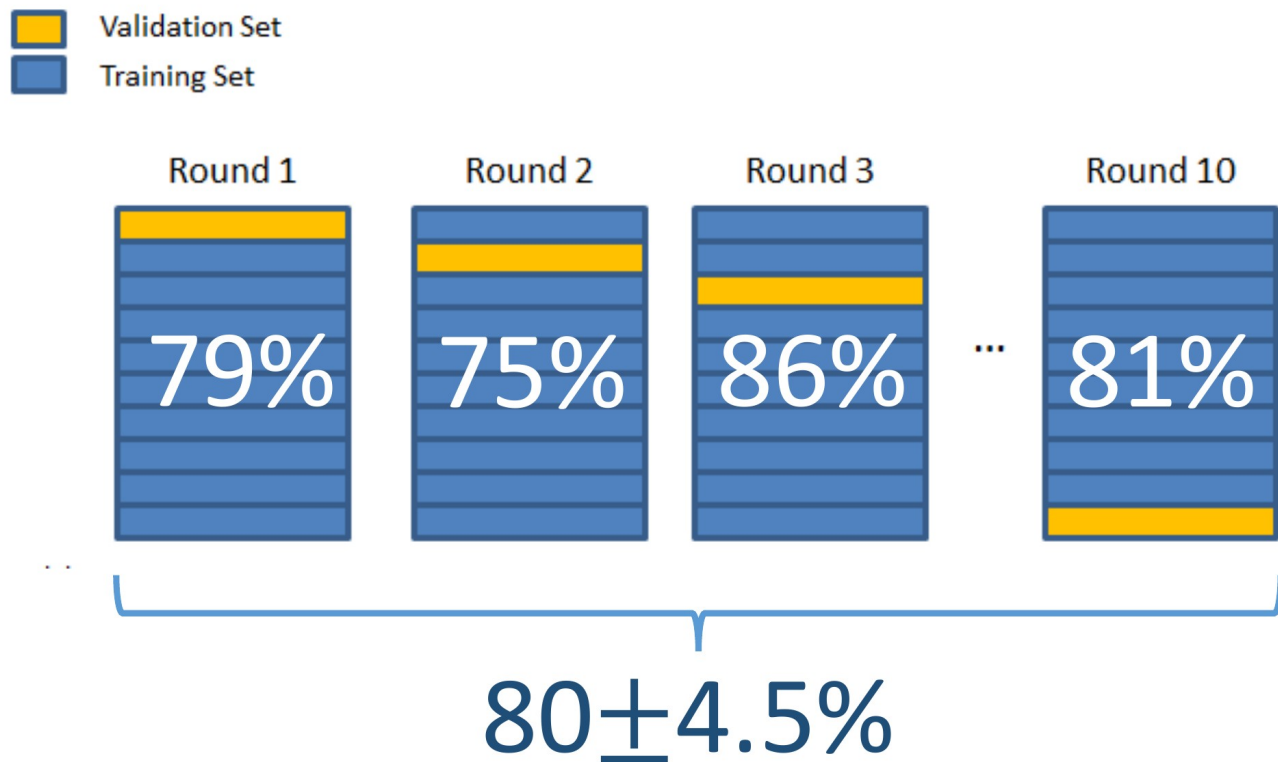
Кросс-валидация — процедура эмпирического оценивания обобщающей способности алгоритмов. С помощью кросс-валидации эмулируется наличие тестовой выборки, которая не участвует в обучении, но для которой известны правильные ответы.

Способы валидации

- Валидация на отложенных данных (Hold-Out)
- Полная кросс-валидация
- Кросс-валидация по отдельным объектам (Leave-One-Out)
- k-fold кросс-валидация
- ...

Стратификация

Учёт разброса и распределения при кросс-валидации



Стабильность решения

Рассматриваем, как меняются настраиваемые параметры модели (зависит от типа модели):

- Стабильные компоненты заслуживают веры
- Если все нестабильно — плохо

Анализ признаков

На одном фолде:

0.211268 Номер
0.147105 Ширина
0.128326 Вес
0.0954617 Параметр 1
0.0688576 Высота
0.057903 Параметр 2
0.0438185 Параметр 3
...

На другом:

0.285714 Номер
0.163265 Параметр 1
0.122449 Высота
0.102041 Параметр 4
0.0816327 Параметр 5
0.0816327 Вес
0.0612245 Параметр 2
...

Анализ зависимости от признаков

Если зависимость от каких-то признаков должна иметь понятный вид, то можно проверить, что модель её находит правильно.

Опасности

- Разные характеристики обучающей и рабочей выборки
- Переобучение на валидации
- Несбалансированные классы

Сложность модели

Какая бывает информация в параметрах:

- про генеральную совокупность
- про выборку
- про random seed

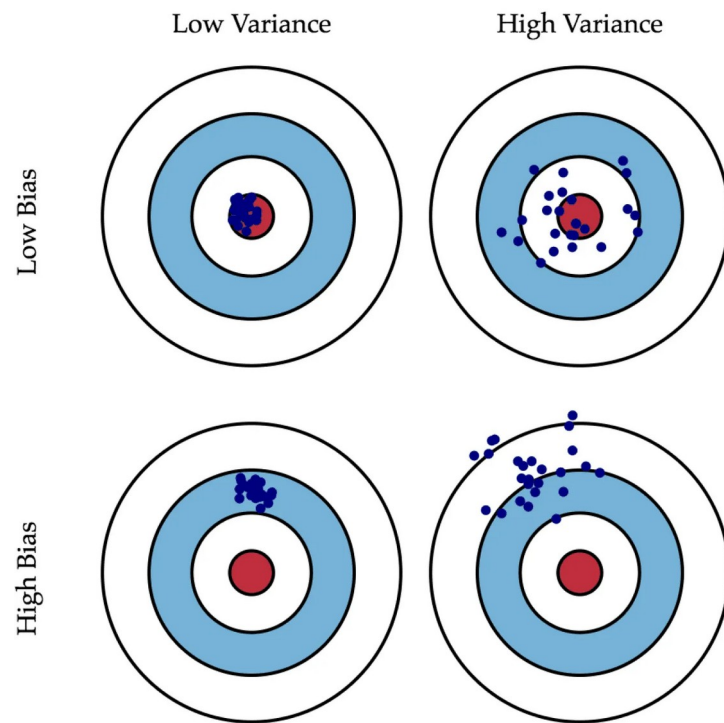
Переобучение и недообучение

- **Переобучение**, переподгонка (overtraining, overfitting) — нежелательное явление, возникающее при решении задач обучения по прецедентам, когда вероятность ошибки обученного алгоритма на объектах тестовой выборки оказывается существенно выше, чем средняя ошибка на обучающей выборке.
- **Недообучение** (underfitting) — нежелательное явление, возникающее при решении задач обучения по прецедентам, когда алгоритм обучения не обеспечивает достаточно малой величины средней ошибки на обучающей выборке.

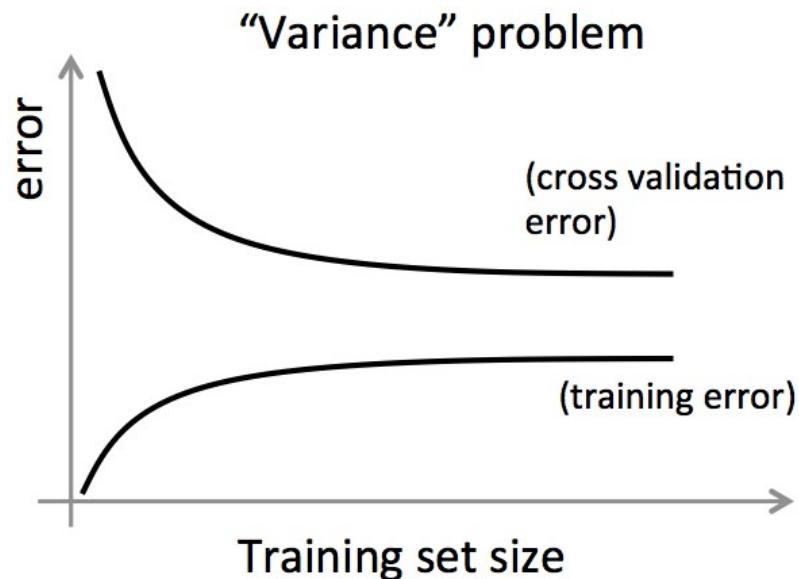
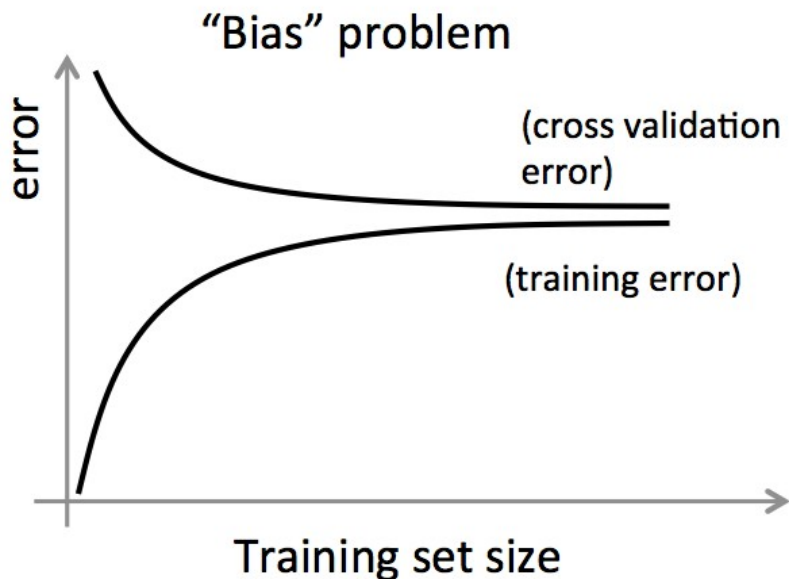
Как можно переобучиться

- Линейные модели: степень полинома
- Деревья решений: глубина дерева
- Нейронные сети: ширина и глубина
- SVM: Kernel trick
- ...

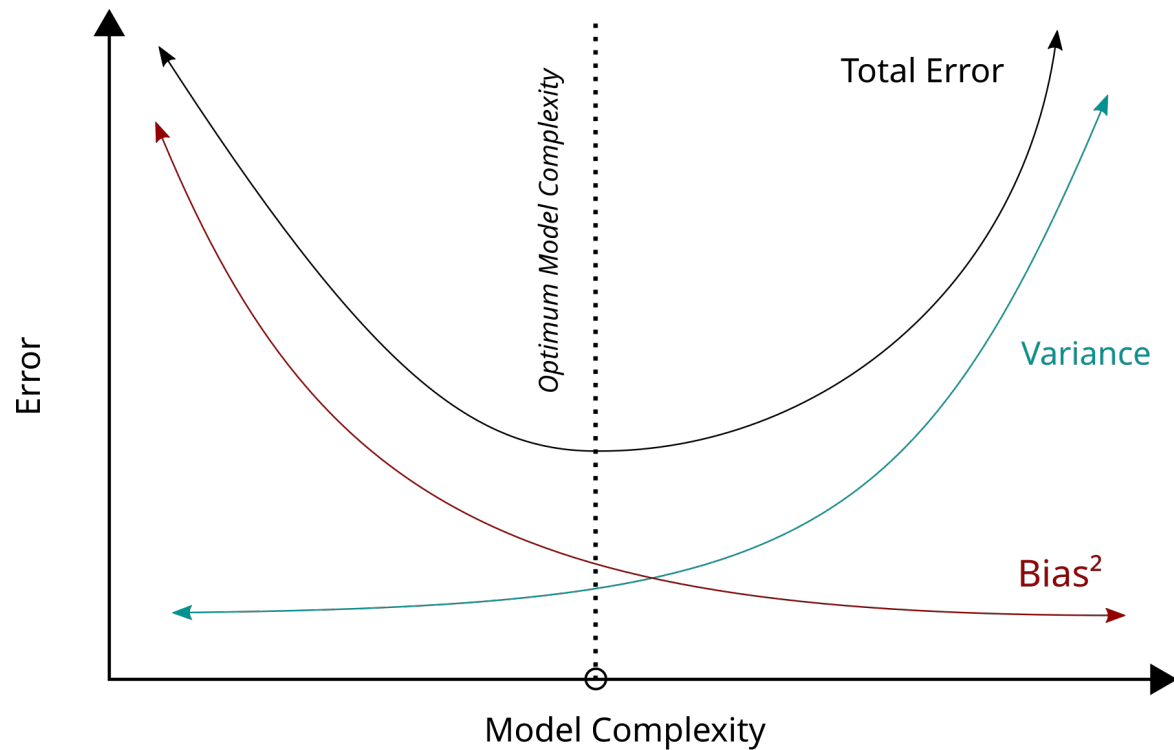
Bias-variance tradeoff



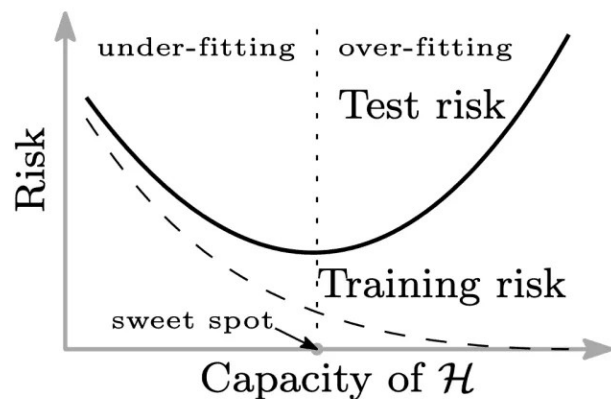
Как понять, где находимся?



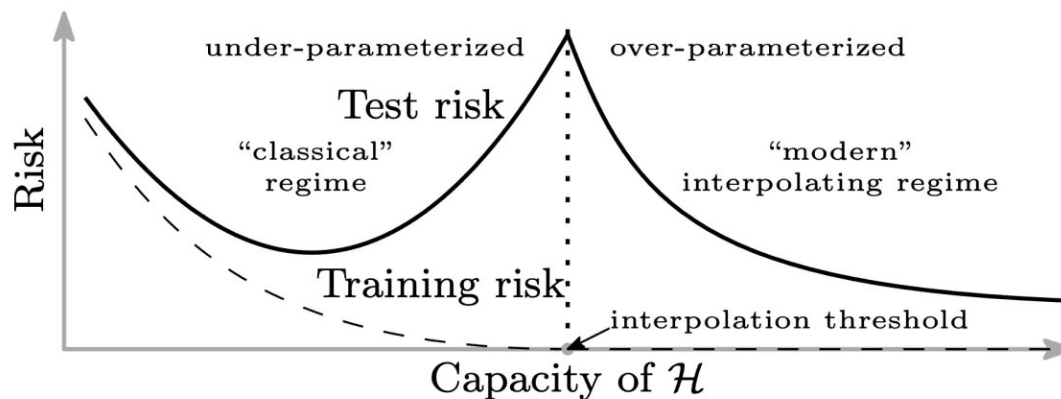
Красивая картинка



Overparametrization



(a)



(b)

КТО ВИНОВАТ И ЧТО ДЕЛАТЬ?

- Увеличение числа примеров для обучения исправляет high variance
- Меньшее число факторов исправляет high variance
- Уменьшение сложности модели исправляет high variance
- Увеличение числа факторов исправляет high bias
- Увеличение сложности модели исправляет high bias