Разбор летучки

Лекция 11

Анализ смещения и разброса

Екатерина Тузова

Постановка задачи

X – множество объектов
Y – ответы в $\mathbb{R}$
Обучающая выборка: $X^l = (x_i, y_i)_{i=1}^l$
Целевая функция: $f : X \to Y$

Набор моделей алгоритмов $A_t : X \to Y$, $t \in T$
Методы обучения $\mu : (X \times Y)^l \to A_t$, $t \in T$

Внешний функционал

$$E_{in} = Q_\mu(X^l) = Q(\mu(X^l), X^l)$$

Внешний функционал по отложенной выборке:

$$E_{out} = Q_\mu(X^t, X^k) = Q(\mu(X^t), X^k)$$

Линейная регрессия

$$E_{in} = Q(w, X^l) = \sum_{i=1}^{l} (f(x_i, w) - y_i)^2 = \|Xw - y\|^2 \to \min_w$$

Внешний функционал по отложенной выборке:

$$E_{out} = Q(w^*, X^k) = \|X^k w^* - y\|^2$$

Постановка задачи

Хотим: $E_{out} \to \min$

Постановка задачи

Хотим: $E_{out} \to \min$

$|A_t| \uparrow \quad \Rightarrow$ больше шансов, что целевая функция f находится во множестве
$|A_t| \downarrow \quad \Rightarrow$ лучше обобщающая способность алгоритма

Постановка задачи

Ошибка алгоритма складывается из двух частей:

- Насколько хорошо A_t может приблизить целевую функцию f
- Можем ли мы на основании X^l выбрать f из A_t

Внутренний функционал качества

Квадратичная функция потерь:

$$L_i(a) = (a(x_i) - f(x_i))^2$$

Внутренний функционал:

$$E_{in} = \sum_{i=1}^{l} (a(x_i, w) - f(x_i))^2 \to \min_w$$

Внешний функционал качества

$$E_{out}(a^{(X^l)}) = \mathbb{E}_x \left[(a^{(X^l)}(x) - f(x))^2 \right]$$

Внешний функционал качества

$$E_{out}(a^{(X^l)}) = \mathbb{E}_x \left[(a^{(X^l)}(x) - f(x))^2 \right]$$

$$\begin{aligned} \mathbb{E}_{(X^l)} \left[E_{out}(a^{(X^l)}) \right] &= \mathbb{E}_{(X^l)} \left[\mathbb{E}_x \left[(a^{(X^l)}(x) - f(x))^2 \right] \right] \\ &= \mathbb{E}_x \left[\mathbb{E}_{(X^l)} \left[(a^{(X^l)}(x) - f(x))^2 \right] \right] \end{aligned}$$

Средний метод

$$\mathbb{E}_{(X^l)}\left[\left(a^{(X^l)}(x) - f(x)\right)^2\right]$$

Средний метод

$$\mathbb{E}_{(X^l)}\left[\left(a^{(X^l)}(x) - f(x)\right)^2\right]$$

$$\bar{a}(x) = \mathbb{E}_{(X^l)}\left[a^{(X^l)}(x)\right] \approx \frac{1}{K}\sum_{k=1}^{K} a^{X_k}(x)$$

$X_1, \ldots, X_K$ – различные обучающие выборки

Средний метод

$$\mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - f(x))^2\right] = \mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - \bar{a}(x) + \bar{a}(x) - f(x))^2\right]$$

Средний метод

$$\mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - f(x))^2\right] = \mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - \bar{a}(x) + \bar{a}(x) - f(x))^2\right]$$

$$= \mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - \bar{a}(x))^2 + (\bar{a}(x) - f(x))^2$$

$$+ 2(a^{(X^l)}(x) - \bar{a}(x))(\bar{a}(x) - f(x))]$$

Средний метод

$$\mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - f(x))^2\right] = \mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - \bar{a}(x) + \bar{a}(x) - f(x))^2\right]$$

$$= \mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - \bar{a}(x))^2 + (\bar{a}(x) - f(x))^2 + 2(a^{(X^l)}(x) - \bar{a}(x))(\bar{a}(x) - f(x))]$$

$$= \mathbb{E}_{(X^l)}\left[(a^{(X^l)}(x) - \bar{a}(x))^2\right] + (\bar{a}(x) - f(x))^2$$

Смещение и разброс

$$\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2] = \underbrace{\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - \bar{a}(x))^2]}_{var(x)} + \underbrace{(\bar{a}(x) - f(x))^2}_{bias(x)}$$

Смещение и разброс

$$\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2] = \underbrace{\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - \bar{a}(x))^2]}_{var(x)} + \underbrace{(\bar{a}(x) - f(x))^2}_{bias(x)}$$

$$\mathbb{E}_{(X^l)}[E_{out}(a^{(X^l)})] = \mathbb{E}_{x}[\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2]]$$

Смещение и разброс

$$\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2] = \underbrace{\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - \bar{a}(x))^2]}_{var(x)} + \underbrace{(\bar{a}(x) - f(x))^2}_{bias(x)}$$

$$\mathbb{E}_{(X^l)}[E_{out}(a^{(X^l)})] = \mathbb{E}_x[\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2]]$$

$$= \mathbb{E}_x[var(x) + bias(x)]$$

Смещение и разброс

$$\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2] = \underbrace{\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - \bar{a}(x))^2]}_{var(x)} + \underbrace{(\bar{a}(x) - f(x))^2}_{bias(x)}$$

$$\mathbb{E}_{(X^l)}[E_{out}(a^{(X^l)})] = \mathbb{E}_x[\mathbb{E}_{(X^l)}[(a^{(X^l)}(x) - f(x))^2]]$$

$$= \mathbb{E}_x[var(x) + bias(x)]$$

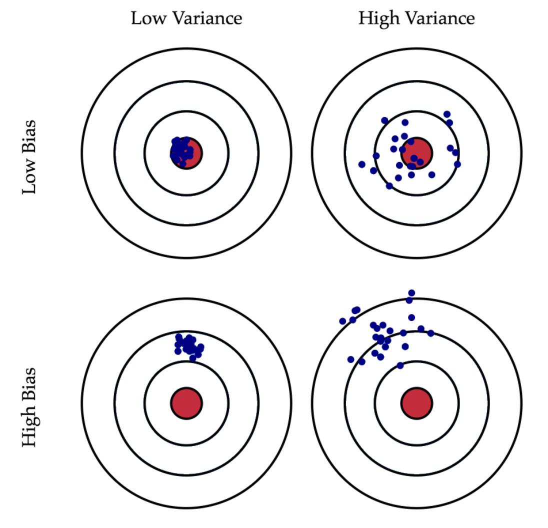
$$= var + bias$$

Смещение и разброс

Смещение (bias) – насколько сложное семейство моделей (средняя ошибка по всевозможным обучающим выборкам X^l)

Разброс (variance) – на сколько чувствителен алгоритм к изменению обучающей выборки (как отличается ошибка, если обучать модель на разных наборах данных)

Смещение и разброс



Смещение и разброс

Как правило, при увеличении сложности модели увеличивается разброс оценки, но уменьшается смещение.

Смещение и разброс

Как правило, при увеличении сложности модели увеличивается разброс оценки, но уменьшается смещение.

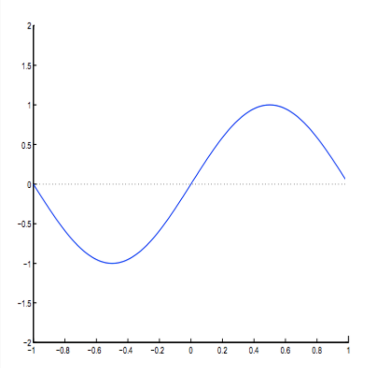
Если же модель слабая, то она не состоянии выучить закономерность, в результате выучивается что-то другое, смещенное относительно правильного решения.

Пример $\sin$

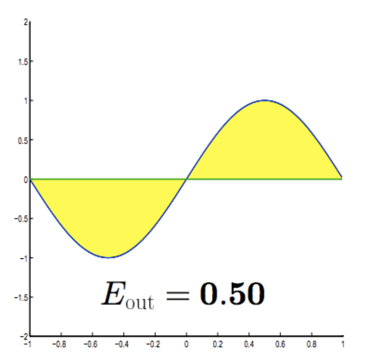
$f(x) = \sin(\pi x)$
Количество объектов $l = 2$

Рассмотрим 2 модели:

$A_1 : h(x) = b$
$A_2 : h(x) = ax + b$

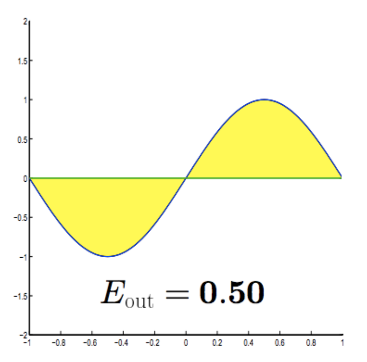


Пример $\sin$

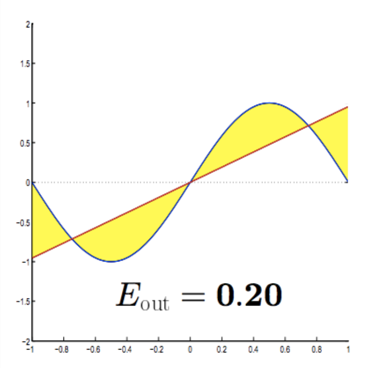


A_1

Пример $\sin$

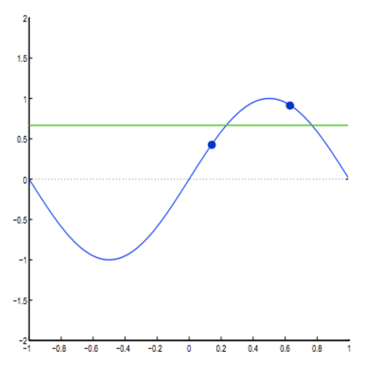


A_1



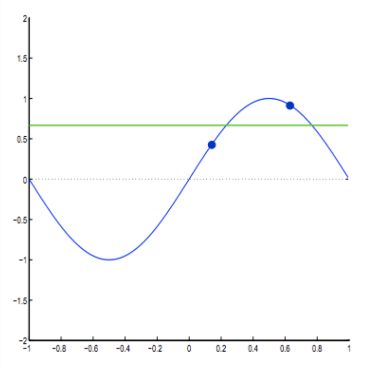
A_2

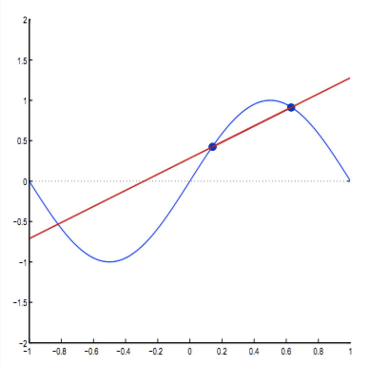
Обучение A_1, A_2



A_1

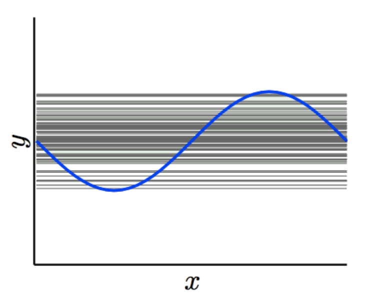
Обучение A_1, A_2



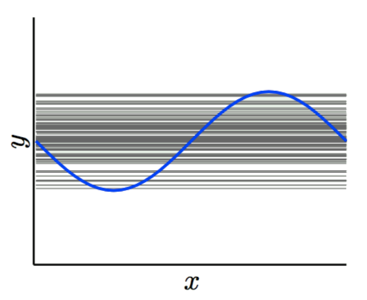


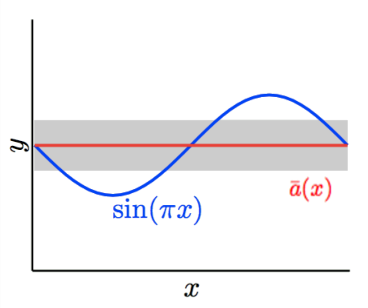
A_1 $\qquad$ A_2

Bias-variance A_1

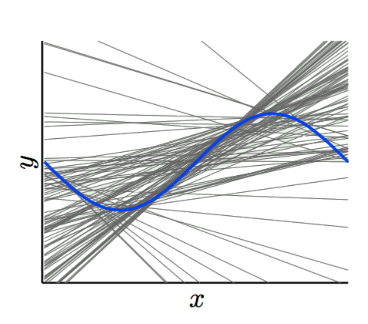


Bias-variance A_1

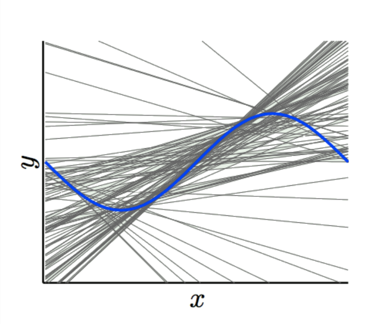


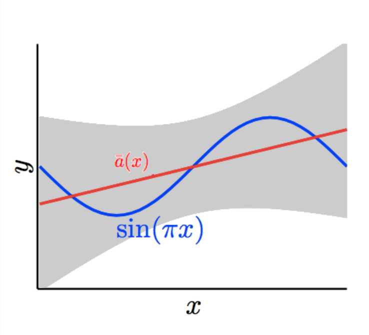


Bias-variance A_2

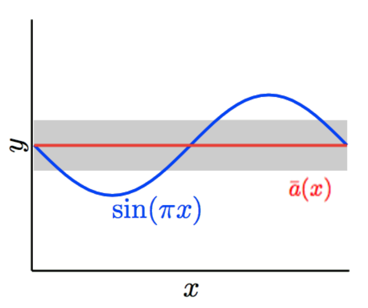


Bias-variance A_2



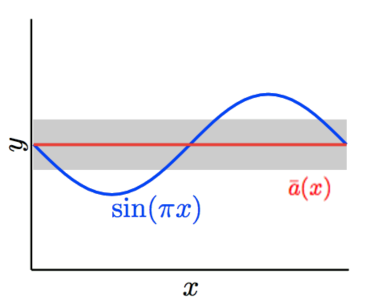


Bias-variance

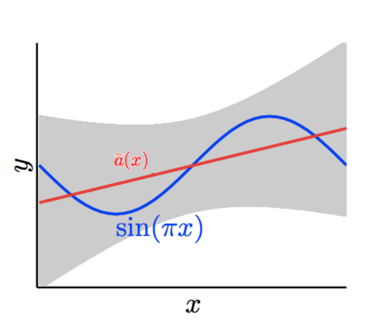


bias = 0.50 var = 0.25

Bias-variance



bias = 0.50 var = 0.25



bias = 0.21 var = 1.69

Мораль

Сложность модели нужно выбирать исходя из доступных данных, а не из предполагаемой сложности целевой функции.

Кривые обучения

Матожидание E_{in} и E_{out}

$$\mathbb{E}_{(X^l)}\left[E_{out}(a^{(X^l)})\right]$$

Матожидание E_{in} и E_{out}

$$\mathbb{E}_{(X^l)}\left[E_{out}(a^{(X^l)})\right]$$

$$\mathbb{E}_{(X^l)}\left[E_{in}(a^{(X^l)})\right]$$

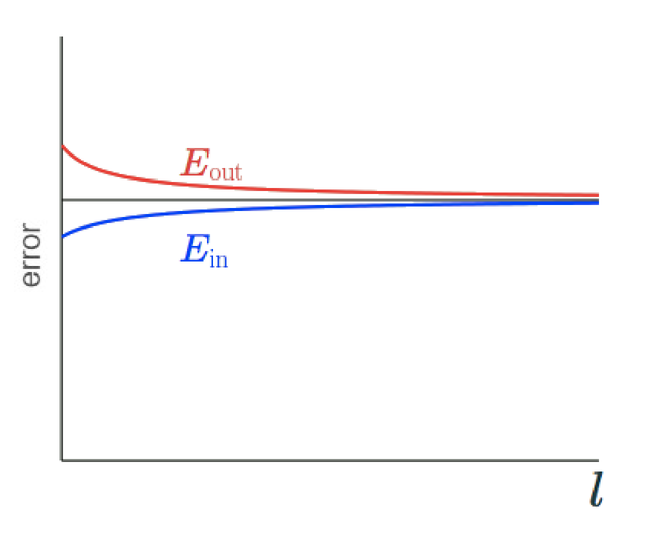
Матожидание E_{in} и E_{out}

$$\mathbb{E}_{(X^l)}\left[E_{out}(a^{(X^l)})\right]$$

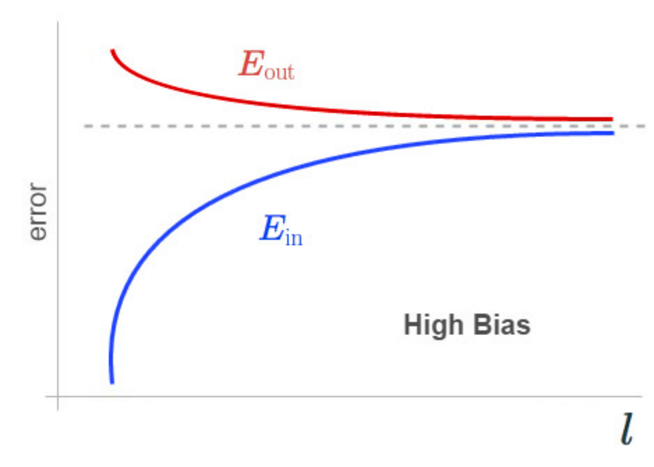
$$\mathbb{E}_{(X^l)}\left[E_{in}(a^{(X^l)})\right]$$

Как зависят от длины выборки?

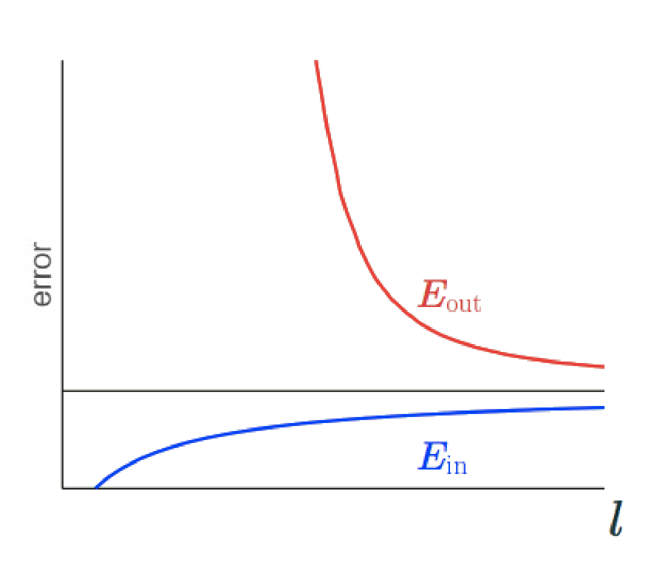
Простая модель



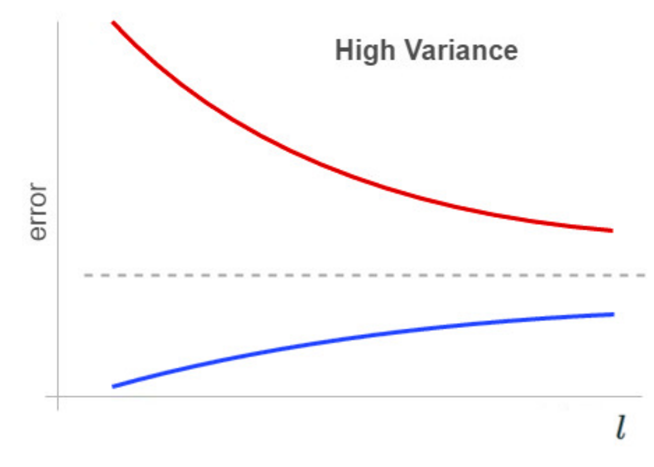
High bias



Сложная модель



High variance



Вопросы?

Что почитать по этой лекции

- Professor Yaser Abu-Mostafa MOOC
- Hastie, T., Tibshirani R. "The Elements of Statistical Learning" Chapter 7

На следующей лекции

- Методы восстановления регрессии
- K neighbors regressor
- Decision tree regression
- Neural Net Regression
- SVR