Теория обучения

3 апреля 2013 г.

Условное математическое ожидание

Определение (Условное м.о.)

Совсем неформальное определение $E(Y|X)$: случайная величина, усредняющая значения Y относительно значений X.
Свойства:
1) Для независимых величин $E(Y|X) = E(Y)$
2) $E(E(Y|X)) = E(Y)$
3) $E(f(X)Y|X) = f(X)E(Y|X)$

Теорема

Пусть есть совместное распределение $P(x, y)$ и некоторая выборка (X, Y). Рассмотрим математическое ожидание ошибки $E(y - f(x))^2$. Лучшая оценка в смысле $E(y - f(x)) \to \min$ будет $E(Y|X)$.

Определение

$$E(y - f(x))^2 = E(y - E(y|x))^2 + E(E(y|x) - f(x))^2$$

$E(y - E(y|x))^2$ - *шум*
$E(E(y|x) - f(x))^2$ - *bias-variance*

Определение

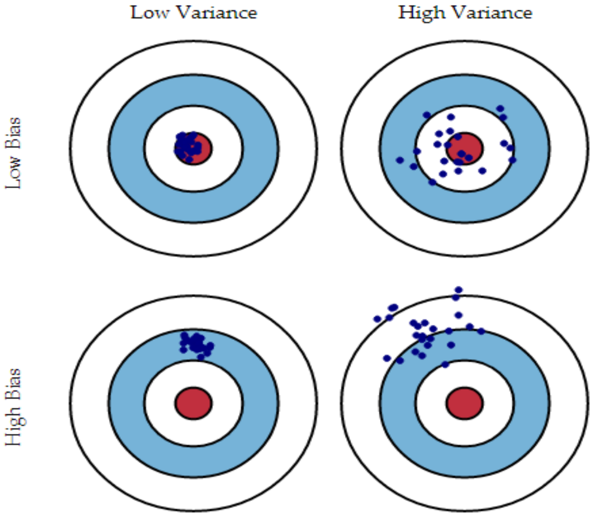
Теперь представим, что есть много датасетов D и оценка зависит от D: $f(x, D)$.

$$E_D(E(y|x) - f(x, D))^2 =$$

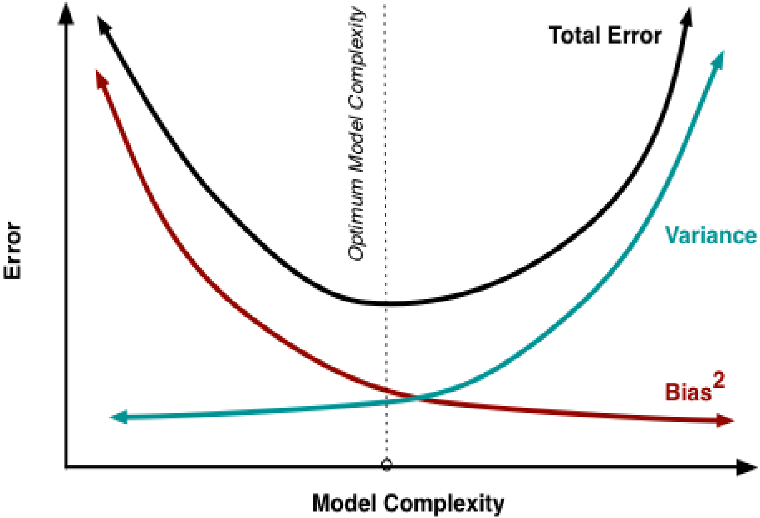
$$= (E(y|x) - E_D(f(x, D)))^2 + E_D(f(x, D) - E_D(f(x, D)))^2$$

$(E(y|x) - E_D(f(x, D)))^2$ - *bias*^{2}
$E_D(f(x, D) - E_D(f(x, D)))^2$ - *variance*

Bias-variance



Bias-variance



Пример

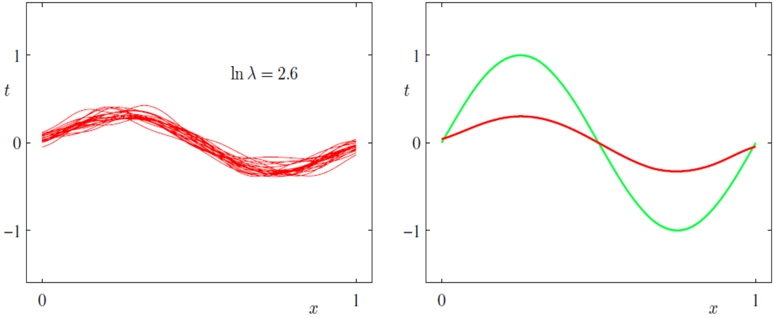
Рассмотрим функцию $h(x) = sin(2\pi x)$. И возьмем $L = 100$ датасетов по $N = 25$ точек. Построим регрессию с регуляризацией L^2 и гауссовскими базисными функциями.

$$\overline{y}(x) = \sum_{l=1}^{L} y^{(l)}(x)$$

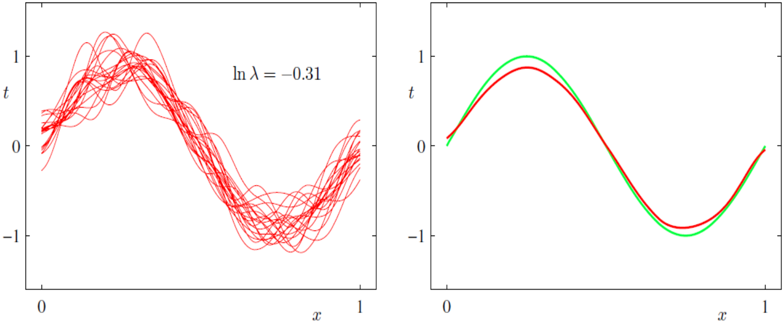
$$bias^2 = \frac{1}{N}\sum_{n=1}^{N}(\overline{y}(x_n) - h(x_n))^2$$

$$\frac{1}{N}\sum_{n=1}^{N}\frac{1}{L}\sum_{l=1}^{L}(y^{(l)}(x_n) - \overline{y}(x_n))^2$$

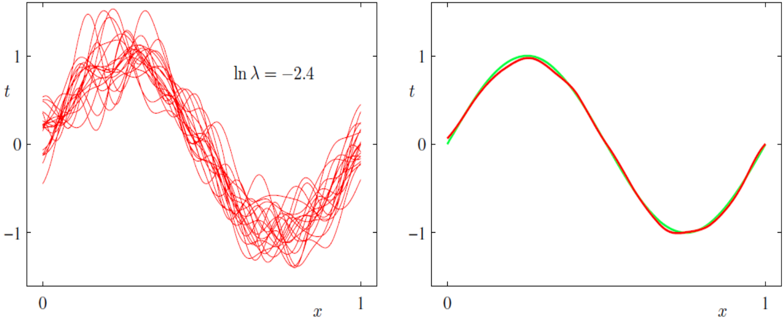
Bias-variance



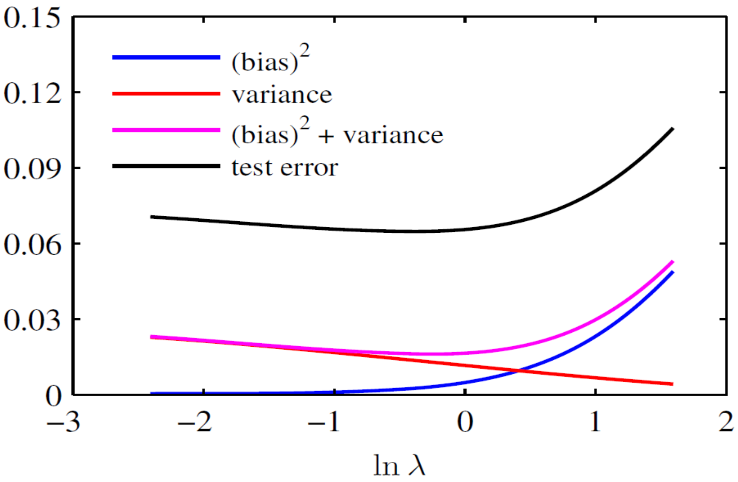
Bias-variance



Bias-variance



Bias-variance



Кросс-валидация

Определение

1) Датасет S разделим на S_{train} и S_cv
2) Тренируем модели на S_{train}
3) Выберем модель с наименьшей ошибкой на S_{cv}
4) (опционально) Тренируем выбранную модель на всем S

К-кратная кросс-валидация

Определение

1) Датасет S *разделим на* k *равных частей:* $S = S_1 \cup \ldots \cup S_k$
2) Для каждого $j = 1 \ldots k$ *тренируем модель* M_i *на* $S_1 \cup \ldots \cup S_{j-1} \cup S_{j+1} \cup \ldots \cup S_k$*, проверяем на* S_j*, считаем ошибки*
3) Выбираем модель с наименьшим средним по j *ошибок*
4) (опционально) Тренируем выбранную модель на всем S