

Смещение и дисперсия (разброс)

Bias and variance

Машинное обучение

- Исходные данные:

\mathbb{X} – множество описаний объектов

\mathbb{Y} – множество допустимых ответов

$f: \mathbb{X} \rightarrow \mathbb{Y}$ – неизвестная целевая зависимость, значения которой известны только на объектах обучающей выборки:

$$X = \{(\vec{x}_1, y_1), \dots, (\vec{x}_l, y_l)\}$$

- Задача обучения по прецедентам:

построить алгоритм (модель) $a: \mathbb{X} \rightarrow \mathbb{Y}$, который приближал бы неизвестную целевую зависимость как на элементах выборки X , так и на всём множестве \mathbb{X}

Машинное обучение

- Функция потерь: $L(y, y_{pred})$ – отклонение ответа модели от правильного ответа на произвольном объекте \vec{x}
- Семейство алгоритмов (моделей): $\mathcal{A} = \{a: \mathbb{X} \rightarrow \mathbb{Y}\}$
– в рамках него ведется поиск оптимального алгоритма
- Эмпирический риск – функционал ошибки, характеризующий среднюю ошибку алгоритма a на выборке X :

$$Q(a, X) = \frac{1}{l} \sum_{i=1}^l L(y, a(\vec{x}_i))$$

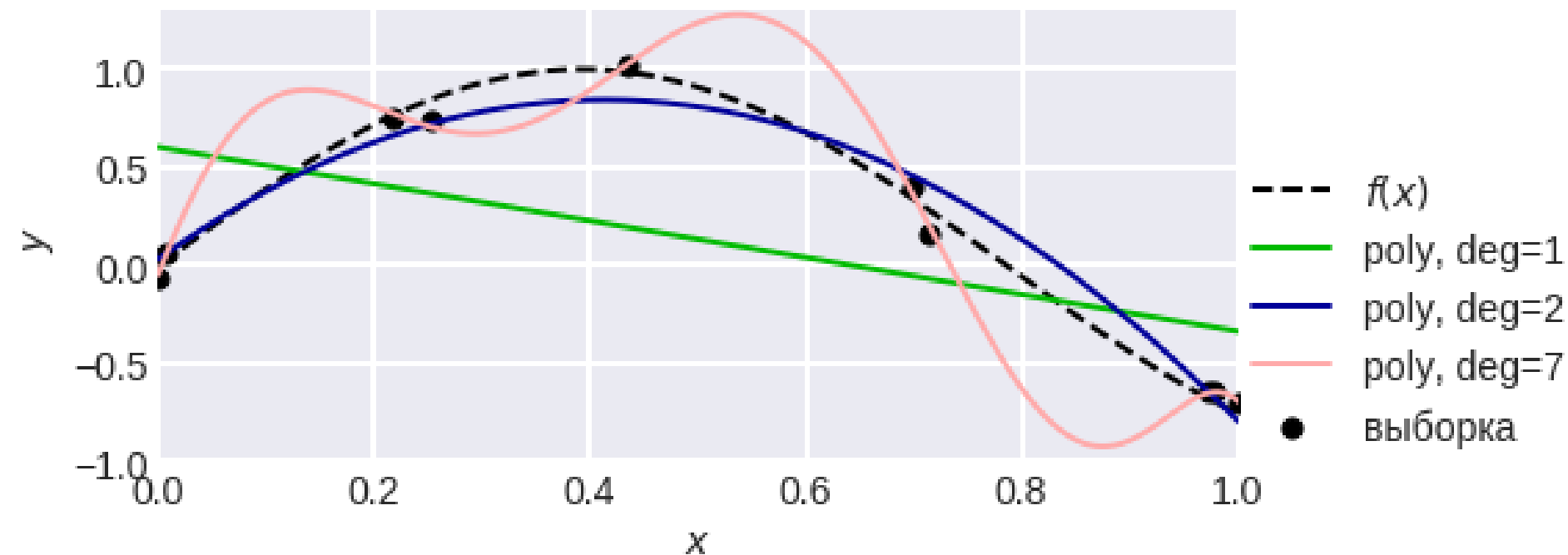
- Метод минимизации эмпирического риска:

$$a = \arg \min_{a \in \mathcal{A}} Q(a, X)$$

Машинное обучение

- Достоинство метода минимизации эмпирического риска: конструктивный и универсальный подход, позволяющий сводить задачу обучения к задачам численной оптимизации
- Недостаток: явление переобучения
 - Если бы не было переобучения, задача машинного обучения сводилась бы к задаче минимизации эмпирического риска
- Сложность семейства алгоритмов – способность алгоритмов семейства настраиваться на обучающую выборку

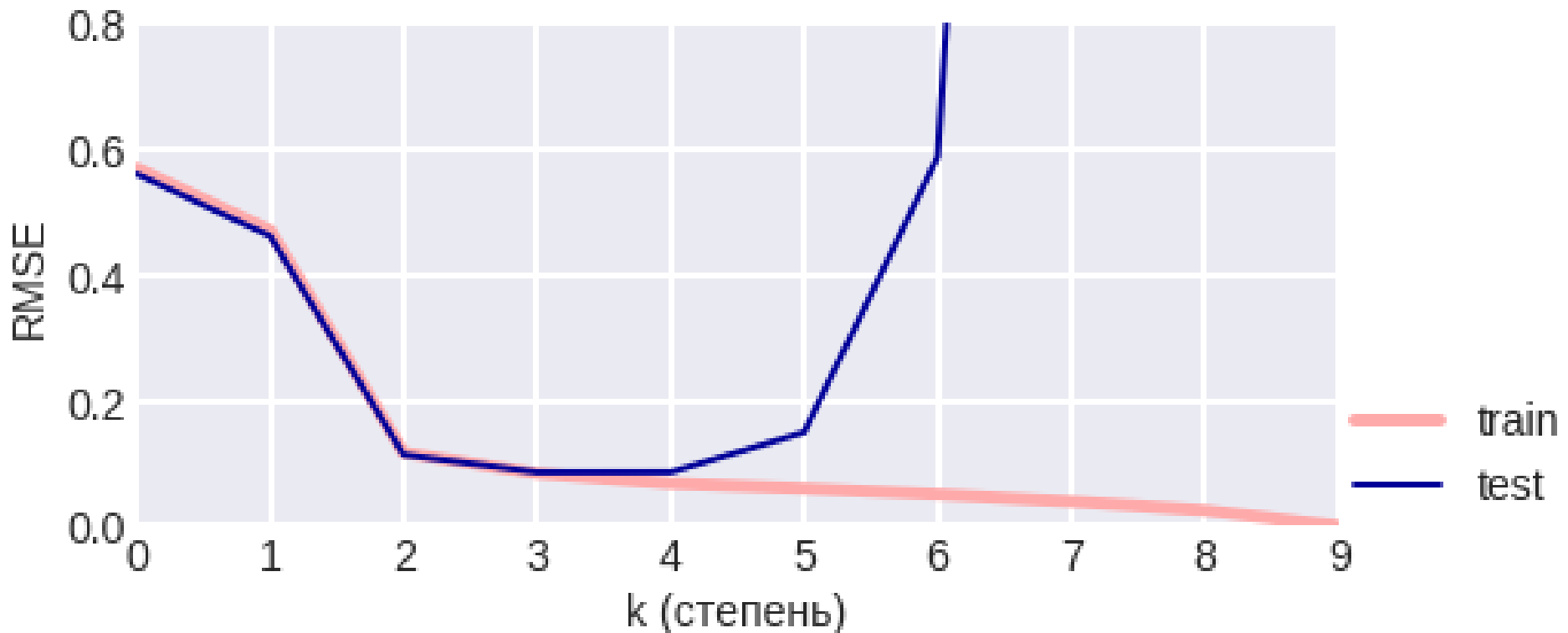
Недообучение и переобучение (underfitting vs. overfitting)



Истинная зависимость: $f(x) = \sin(x) + \text{шум}$

Недообучение и переобучение

- Зависимость ошибки на обучении и тесте от степени полинома



Мат. ожидание и дисперсия

- Математическое ожидание (M) или expectation (E) – среднее значение случайной величины при стремлении количества её измерений к бесконечности. Для дискретного распределения:

$$\mu = M[X] = \sum_{i=1}^{\infty} x_i p_i$$

- Дисперсия (Variance, Var) – математическое ожидание квадрата отклонения случайной величины от её математического ожидания:

$$D[X] = M[(X - M[X])^2] = M[X^2] - (M[X])^2$$

$$D[X] = \sum_{i=1}^{\infty} p_i (x_i - M[X])^2$$

- Среднеквадратическое (стандартное) отклонение (standard deviation, std.dev.):

$$\sigma = \sqrt{D[X]}, \quad D[X] = \sigma^2$$

Истинная зависимость

- Целевая (истинная) зависимость:

$$y(x) = f(x) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2) - \text{шум}$$

$$y(x) = f(x) + \mathcal{N}(0, \sigma^2)$$

$$M[y(x)] = M[f(x)] = f(x)$$

(зависимость детерминирована,

матожидание линейно: $M[aX + bY] = aM[X] + bM[y]$)

$$D[y(x)] = \sigma^2$$

Смещение и дисперсия

- Математическое ожидание квадратичной функции потерь в одной точке:

$$\begin{aligned} M[(y - a)^2] &= M[(y^2 + a^2 - 2ya)] = \\ &= M[y^2] + M[a^2] - 2M[ya] = \\ &= M[y^2] - (M[y])^2 + (M[y])^2 + M[a^2] - (M[a])^2 + (M[a])^2 - 2M[ya] = \\ &= \textcolor{red}{M[y^2]} - (\textcolor{red}{M[y]})^2 + (M[y])^2 + \textcolor{green}{M[a^2]} - (\textcolor{green}{M[a]})^2 + (M[a])^2 - 2M[ya] = \\ &= \textcolor{red}{D[y]} + \textcolor{green}{D[a]} + (M[y])^2 + (M[a])^2 - 2M[ya] = \\ &= \textcolor{red}{D[y]} + \textcolor{green}{D[a]} + f^2 + (M[a])^2 - 2fM[a] = \\ &= \textcolor{red}{D[y]} + \textcolor{green}{D[a]} + (M[f - a])^2 = \\ &= \textcolor{red}{\sigma^2} + \textcolor{green}{variance(a)} + \textcolor{purple}{bias^2(f, a)} \end{aligned}$$

Смещение и дисперсия

$$M[(y - a)^2] = \sigma^2 + \textit{variance}(a) + \textit{bias}^2(f, a)$$

– ошибка раскладывается на три составляющие:

- Первая составляющая σ^2 характеризует шум в данных и равна ошибке идеальной модели $f(x)$
 - Невозможно построить модель, имеющую меньшую среднеквадратичную ошибку
- Вторая компонента $\textit{variance}(a)$ характеризует дисперсию, то есть разброс ответов обученных моделей относительно среднего ответа
 - Дисперсия характеризует разнообразие алгоритмов (из-за случайности обучающей выборки, в том числе шума, и стохастической природы настройки)

Смещение и дисперсия

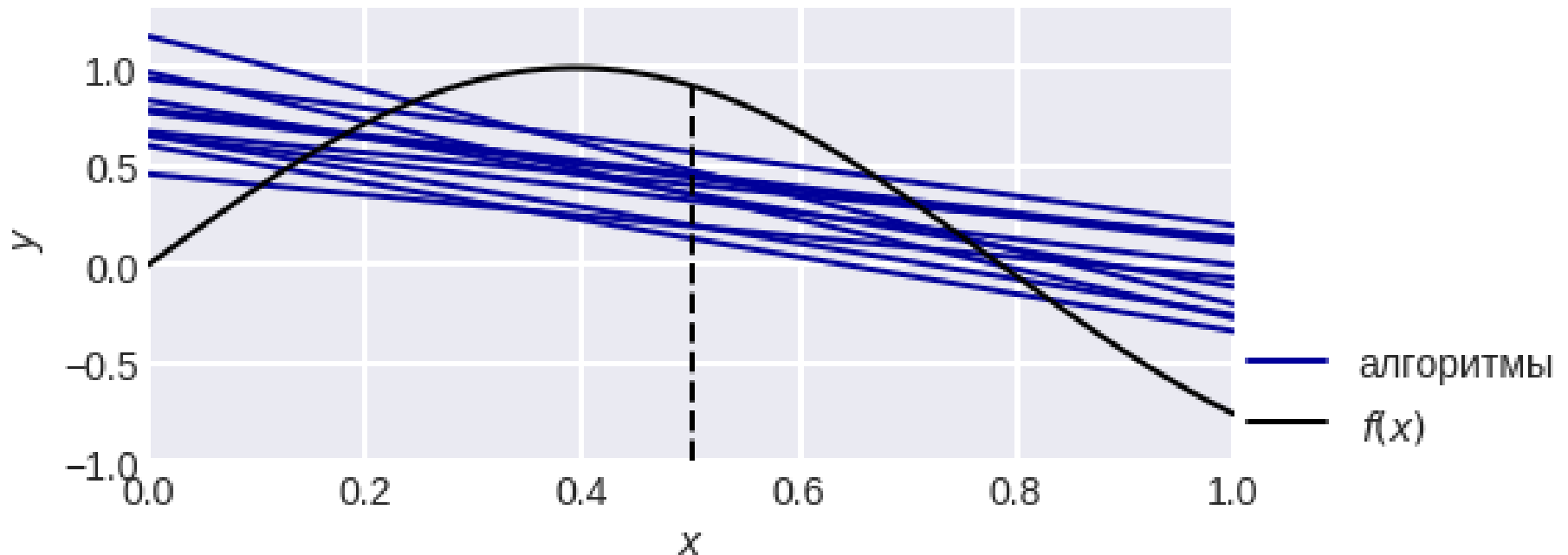
$$M[(y - a)^2] = \sigma^2 + \text{variance}(a) + \text{bias}^2(f, a)$$

– ошибка раскладывается на три составляющие:

- Третья компонента $\text{bias}^2(f, a)$ характеризует смещение модели, то есть отклонение среднего ответа обученной модели от ответа идеальной модели
 - Смещение характеризует способность модели настраиваться на целевую зависимость

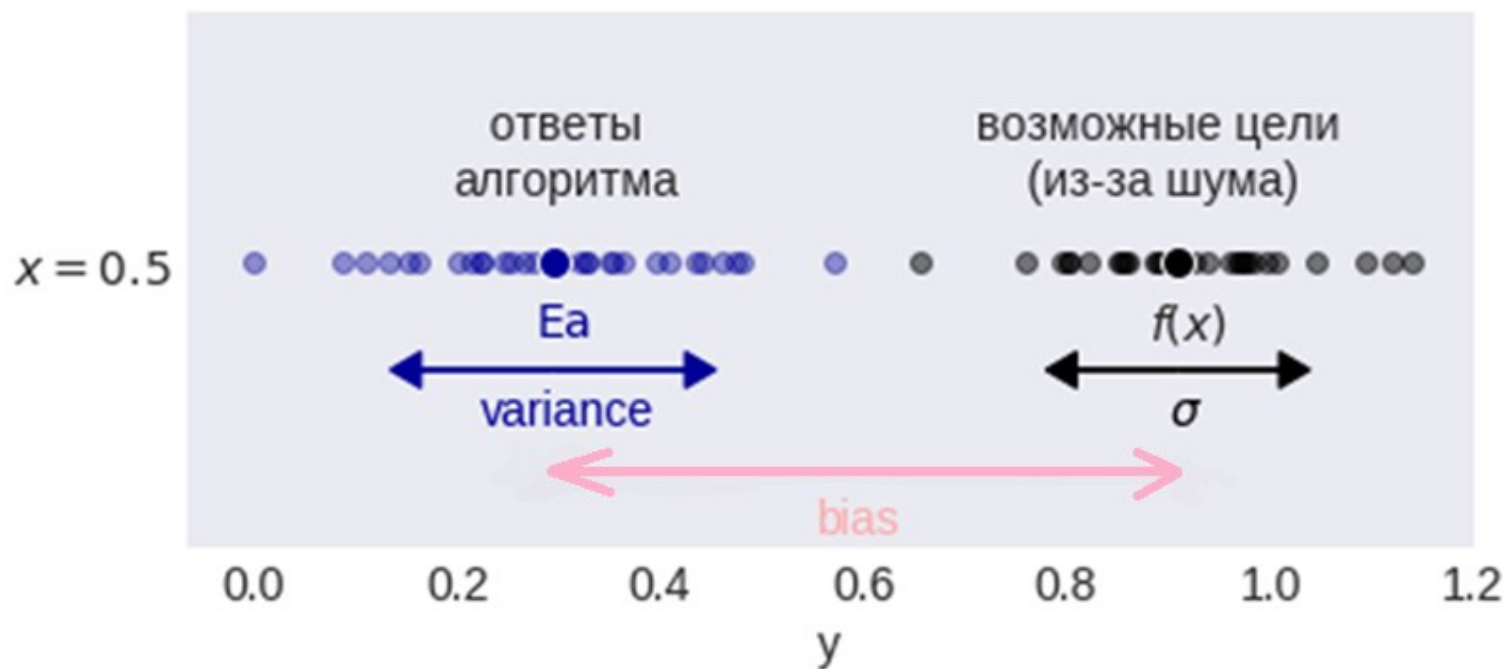
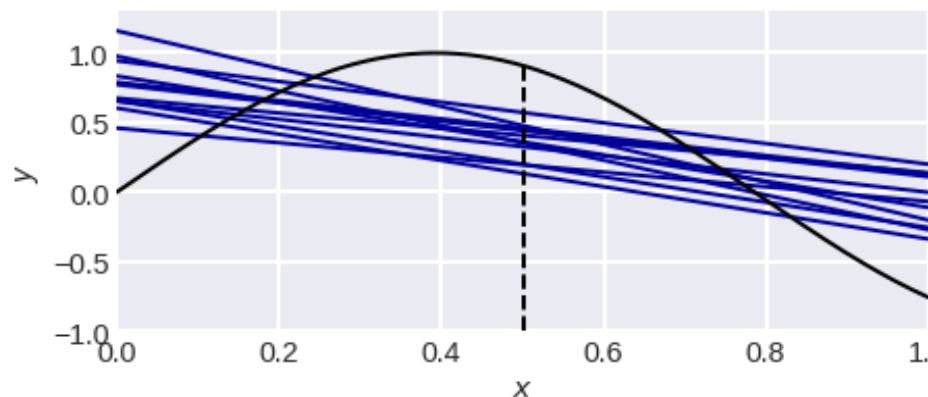
Смещение и дисперсия

- Полиномы 1-й степени, настроенные на разных обучающих выборках:



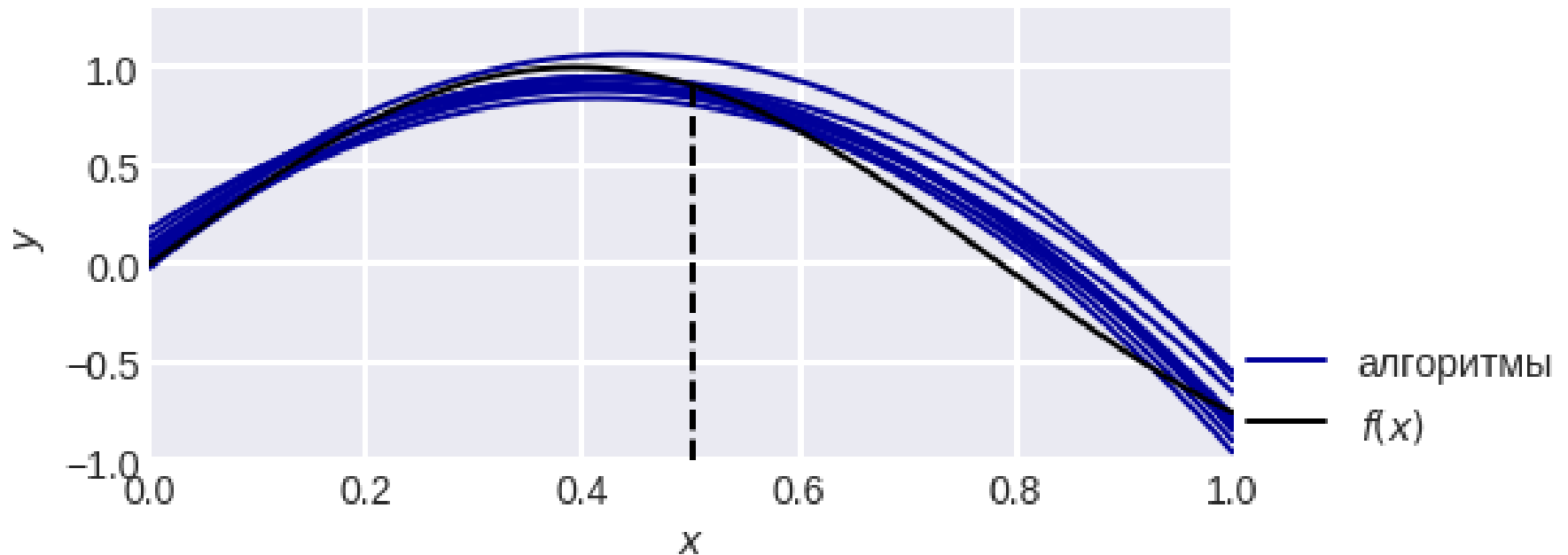
Смещение и дисперсия

- Шум, разброс и смещение:



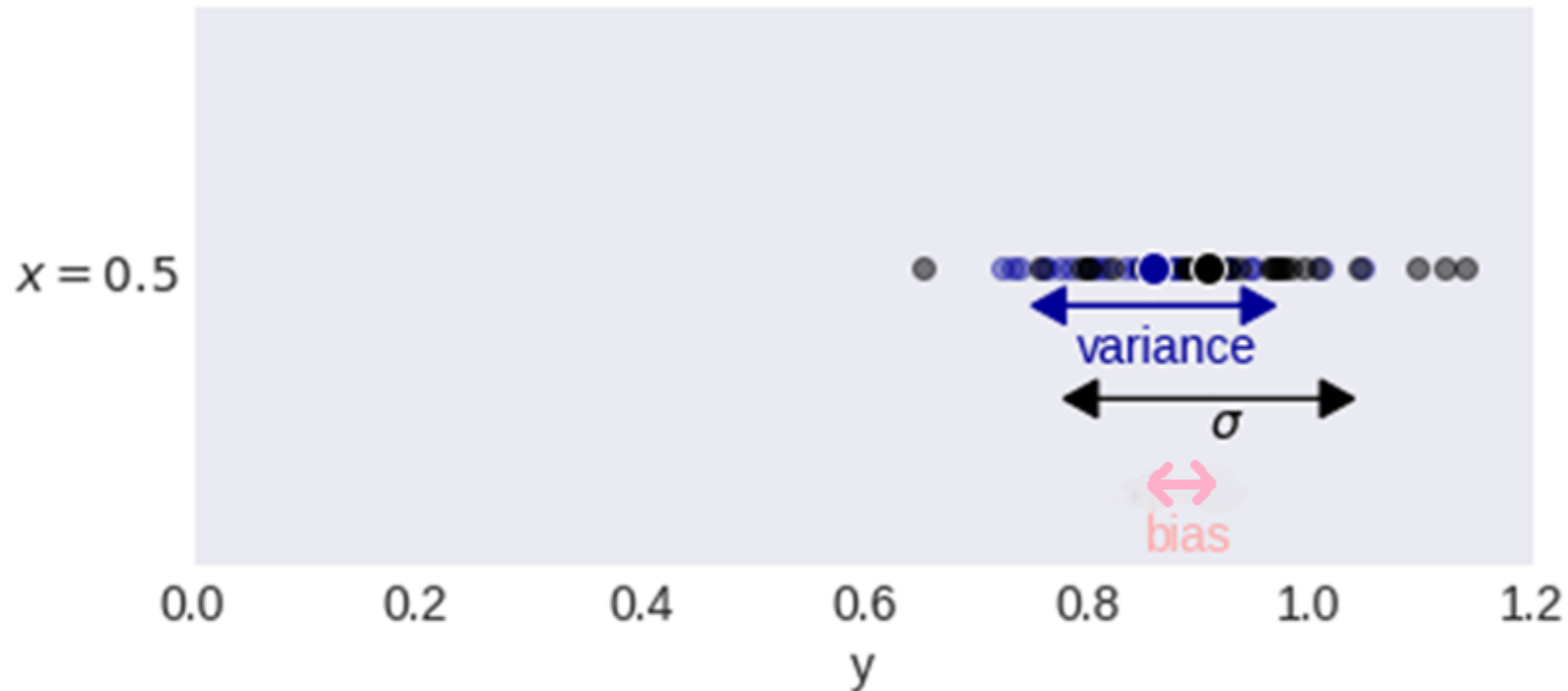
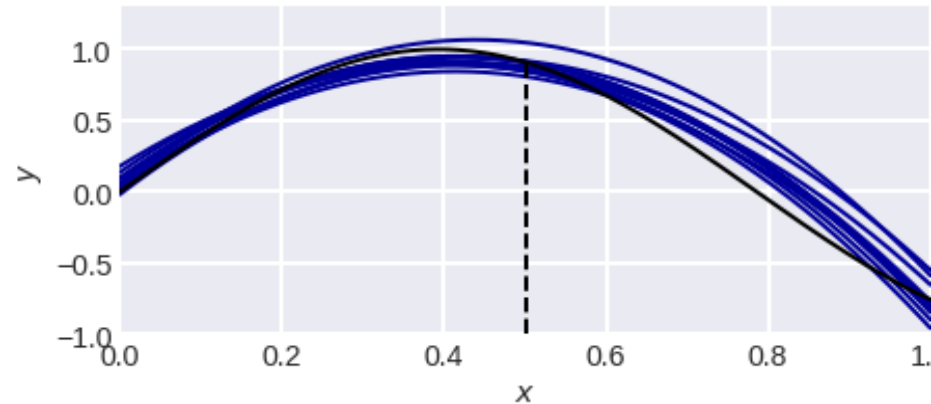
Смещение и дисперсия

- Полиномы 2-й степени, настроенные на разных обучающих выборках:



Смещение и дисперсия

- Шум, разброс и смещение:

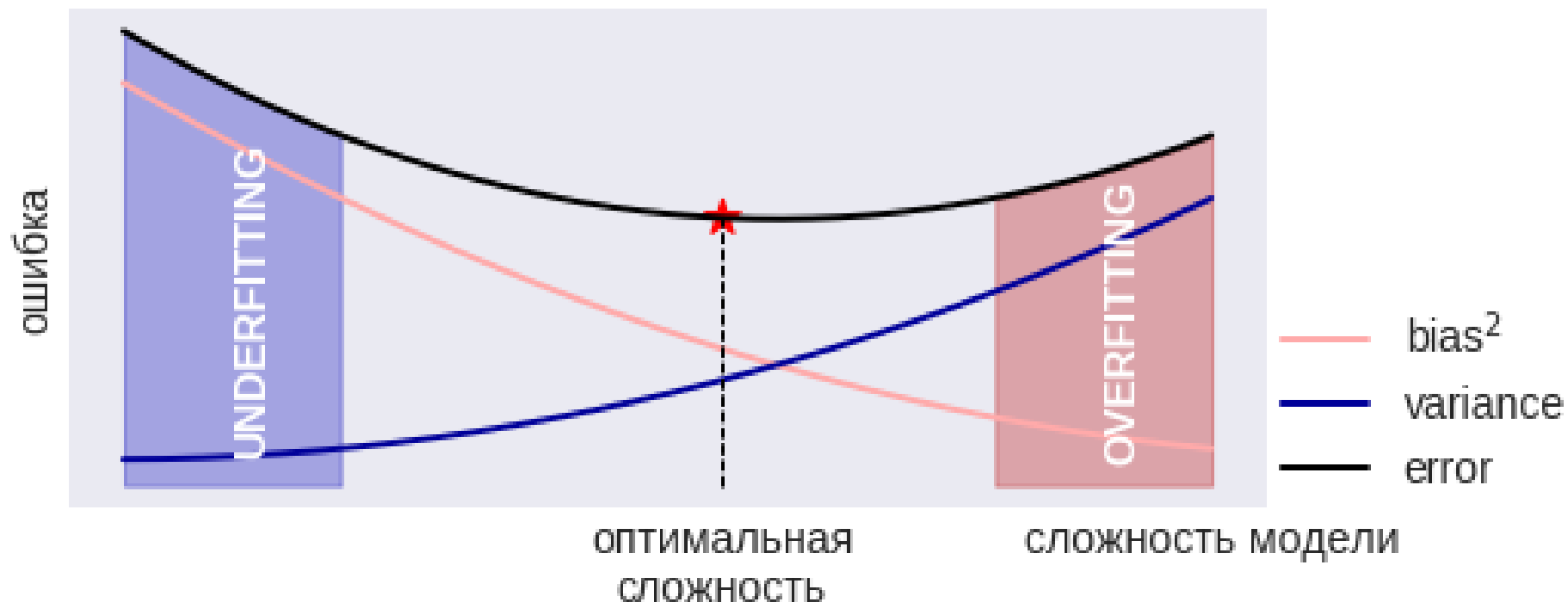


Смещение и дисперсия

- Дартс:



Смещение и дисперсия (bias-variance tradeoff)



Литература

- Дьяконов А. Смещение (bias) и разброс (variance)
 - <https://dyakonov.org/2018/04/25/смещение-bias-и-разброс-variance-модели-алгорит/>