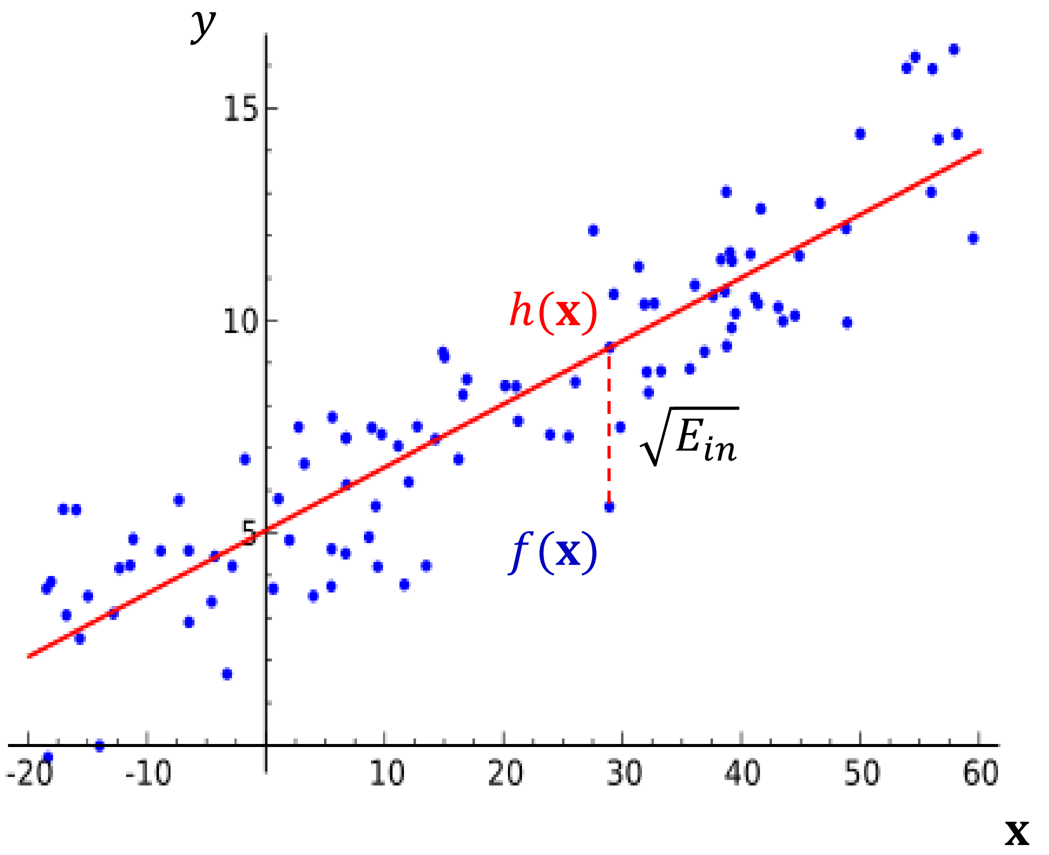
Регрессия

Линейная регрессия

$$E_{in}(h, \mathbf{x}) = \big(h(\mathbf{x}) - f(\mathbf{x})\big)^2$$

$$E_{in}(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}_i - y_i\right)^2 = \frac{1}{N}\left|\left|\mathbf{X}\mathbf{w} - \mathbf{y}\right|\right|_2^2$$

$$\mathbf{X} = \begin{bmatrix} - & \mathbf{x}_1^{\mathrm{T}} & - \\ - & \mathbf{x}_2^{\mathrm{T}} & - \\ & \vdots & \\ - & \mathbf{x}_N^{\mathrm{T}} & - \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$



Линейная регрессия

$$E_{in}(\mathbf{w}) = \frac{1}{N}\left||\mathbf{Xw} - \mathbf{y}|\right|_2^2$$

$$\nabla E_{in}(\mathbf{w}) = \frac{2}{N}\mathbf{X}^{\mathrm{T}}(\mathbf{Xw} - \mathbf{y}) = 0$$

$$\mathbf{X}^{\mathrm{T}}\mathbf{Xw} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$$

$$\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}, \qquad \mathbf{w} = \mathbf{X}^{\dagger}\mathbf{y}$$

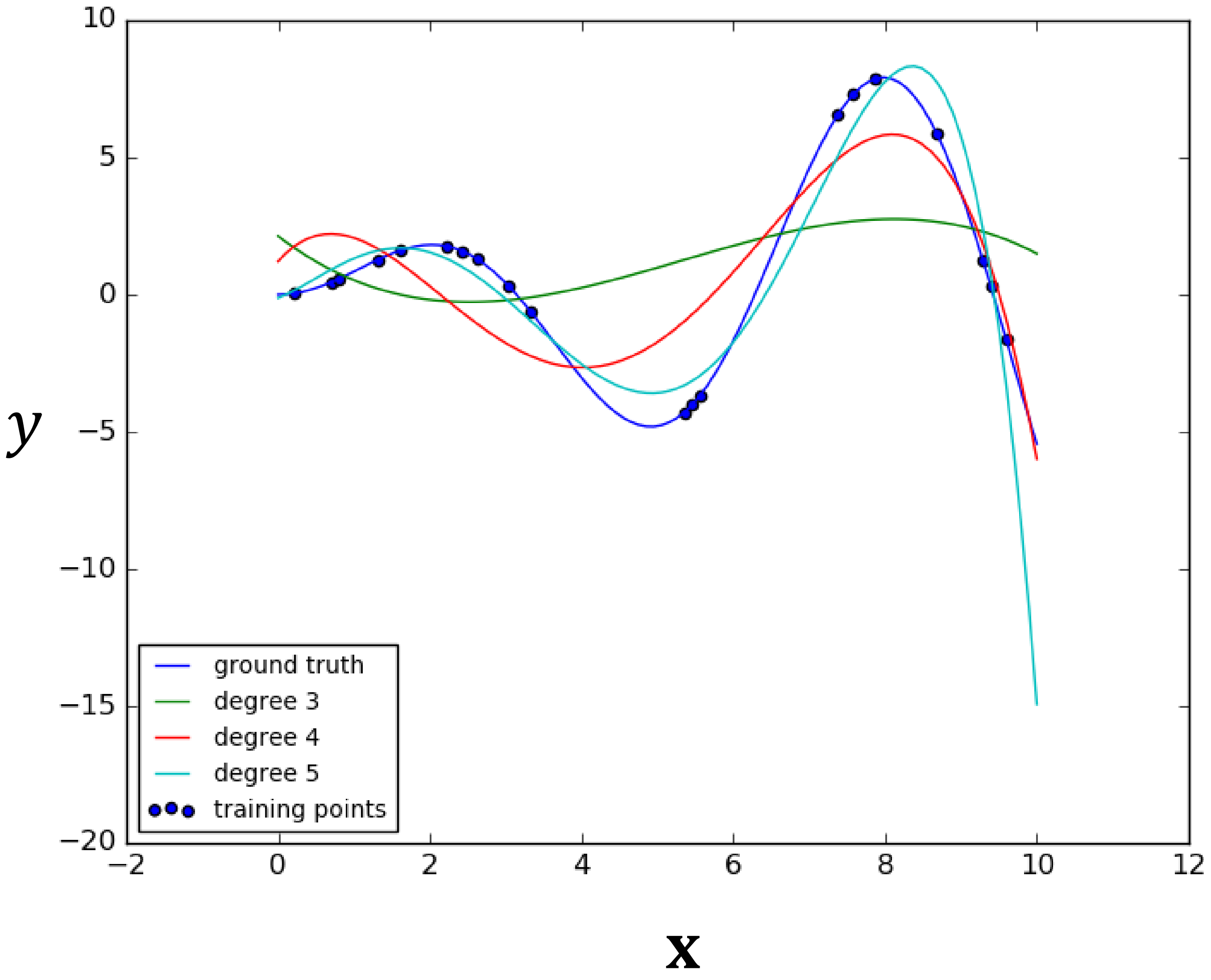
Полиномиальная регрессия

$$X \to Z$$

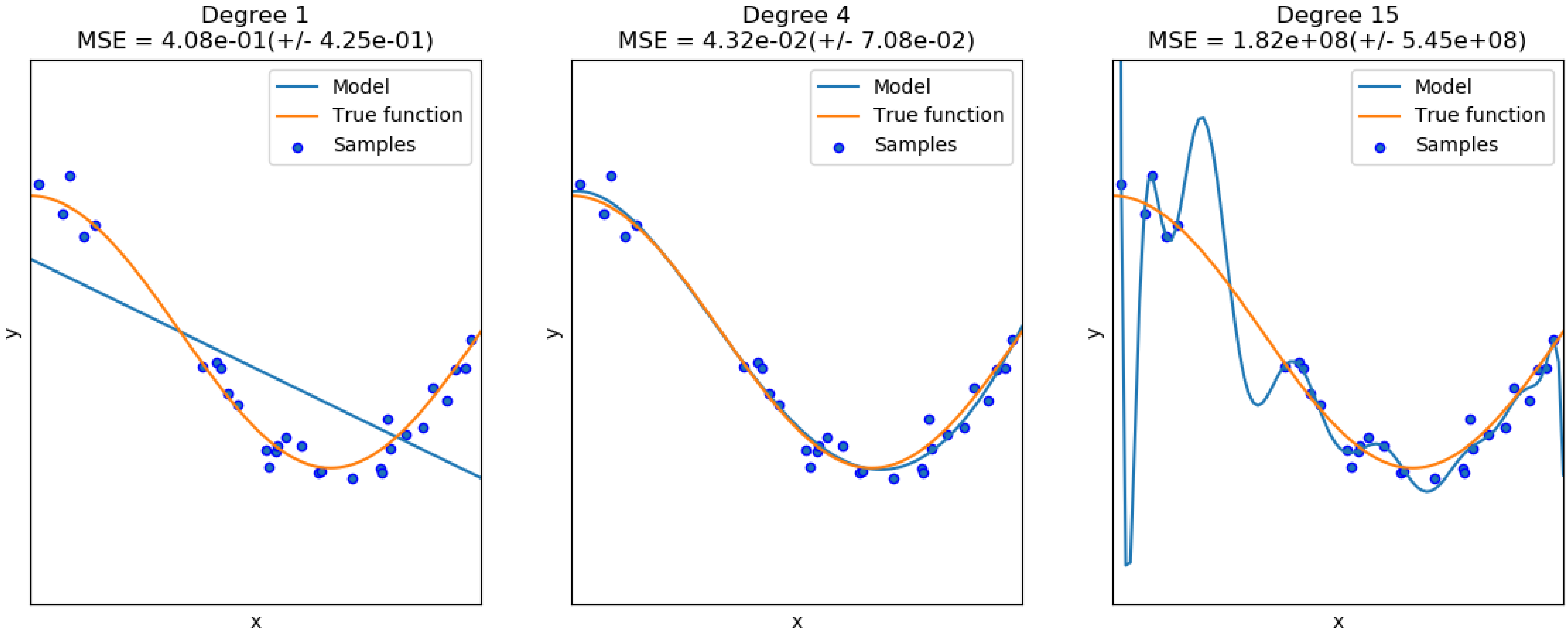
$$x \to [1, x, x^2]$$

$$[x_1, x_2] \to [1, x_1, x_2, x_1^2, x_2^2, x_1 x_2]$$

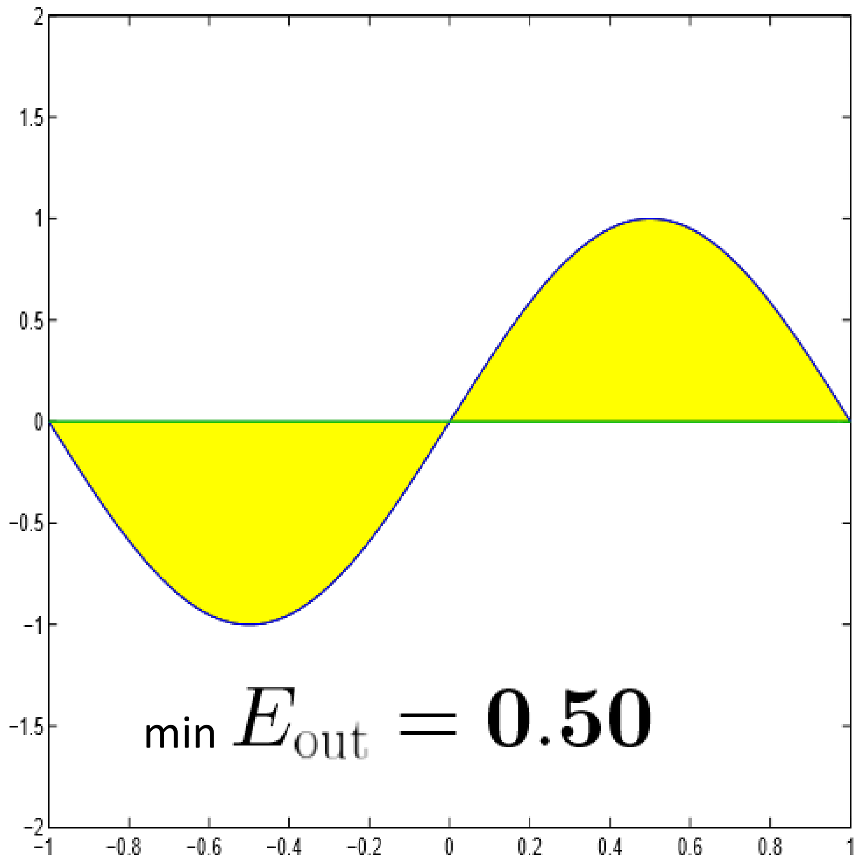
etc...

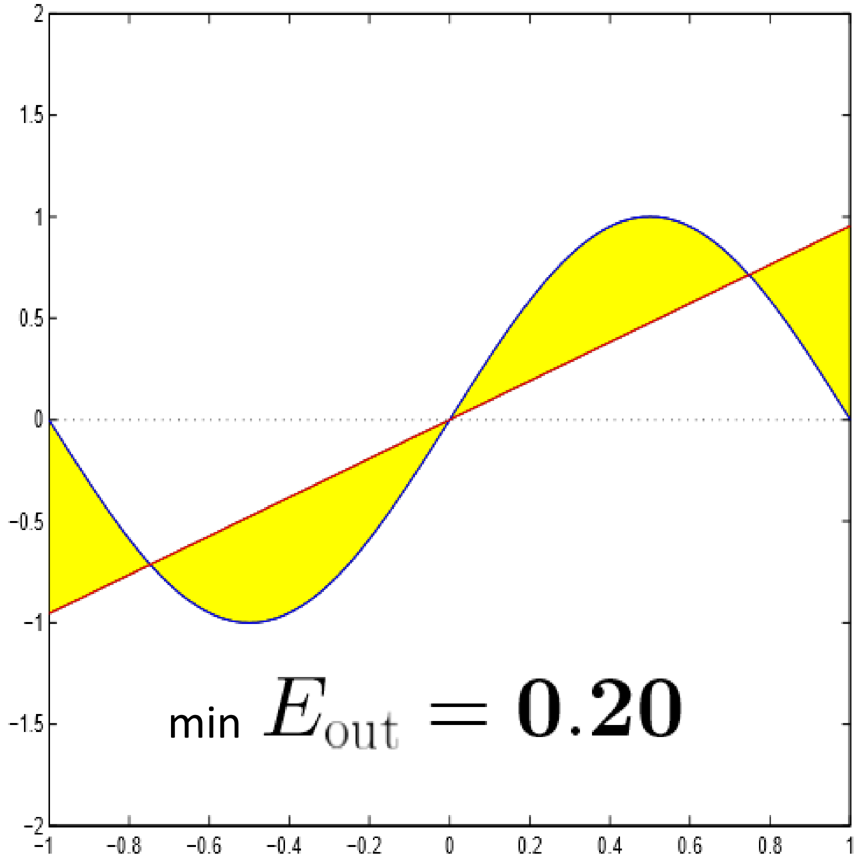


Полиномиальная регрессия

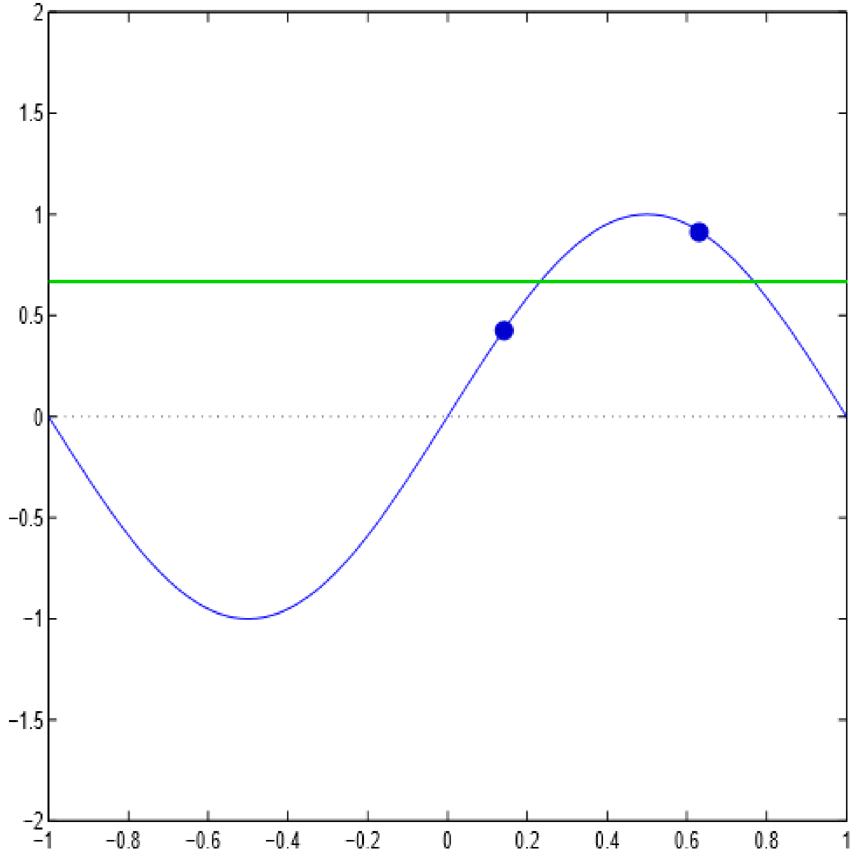


Цель - синус

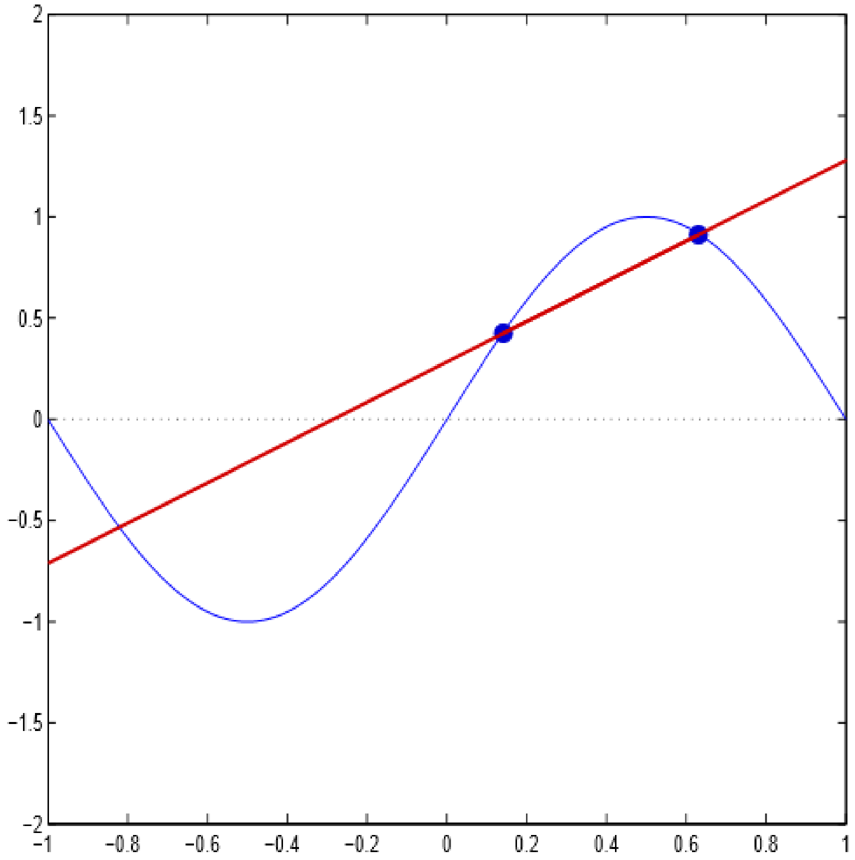




Цель - синус

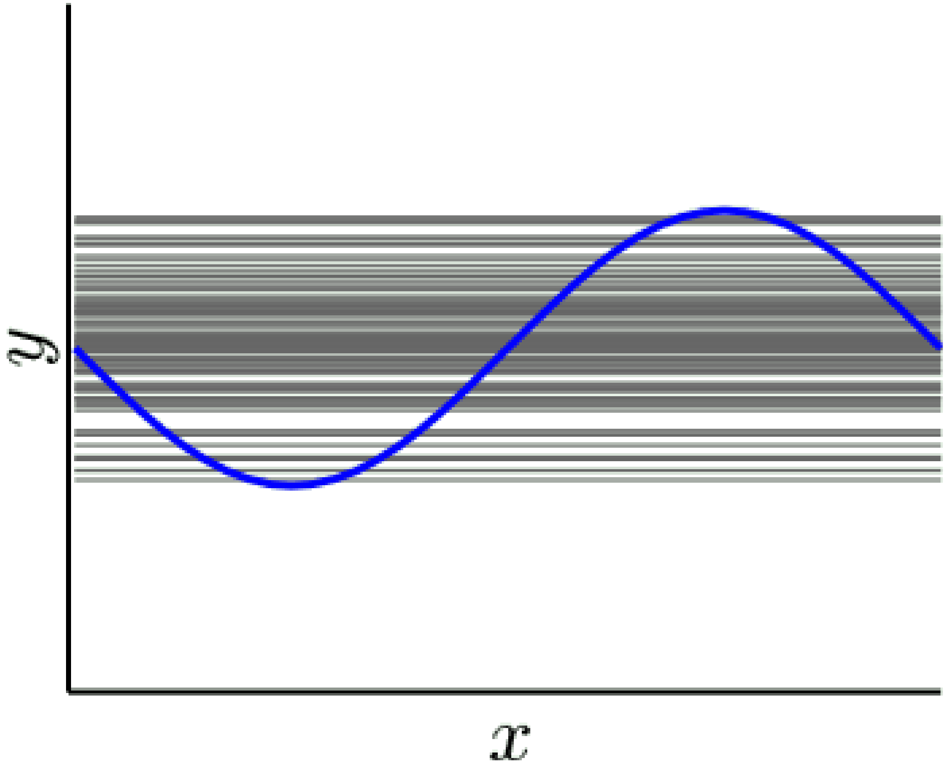


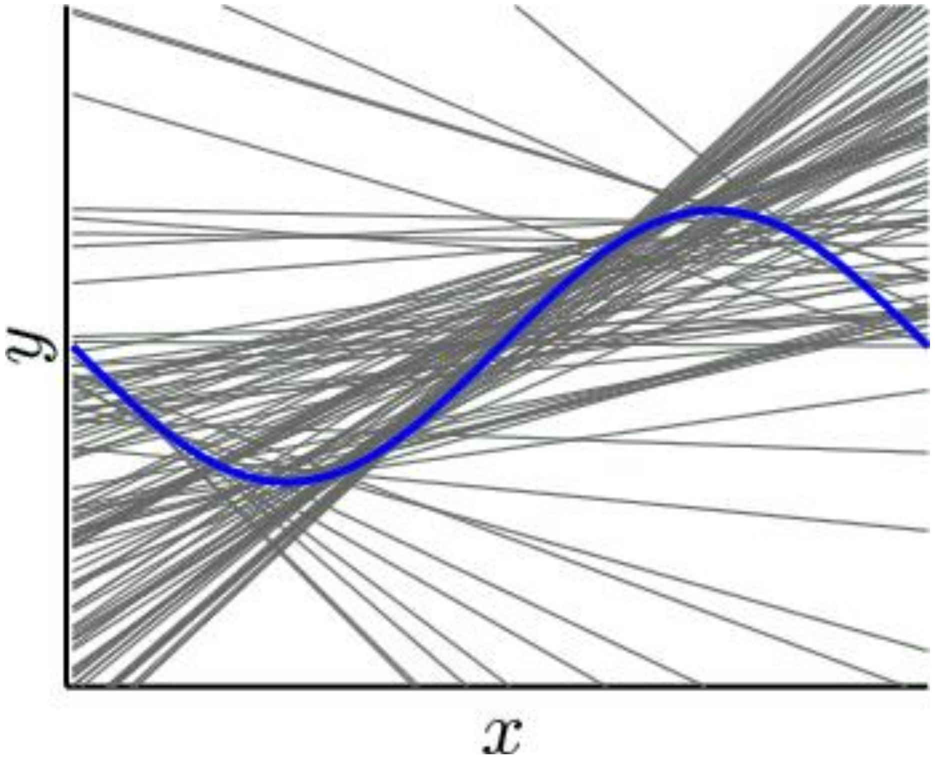
2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1



2
1.5
1
0.5
0
−0.5
−1
−1.5
−2
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1

Цель - синус





Bias and Variance

$$E_{out}(h^D) = \mathbb{E}_{\mathbf{x}}\left[\left(h^D(\mathbf{x}) - f(\mathbf{x})\right)^2\right] \; (D \; - \text{датасет})$$

$$\mathbb{E}_D[E_{out}(h^D)] = \mathbb{E}_D\left[\mathbb{E}_{\mathbf{x}}\left[\left(h^D(\mathbf{x}) - f(\mathbf{x})\right)^2\right]\right] = \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - f(\mathbf{x})\right)^2\right]\right]$$

$$\bar{h}(\mathbf{x}) = \mathbb{E}_D[h^D(\mathbf{x})] \; (\text{средняя гипотеза})$$

$$\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - f(\mathbf{x})\right)^2\right] = \mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x}) + \bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2\right] =$$

$$= \mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)^2 + \left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2 + 2\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)\left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)\right] =$$

$$= \mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)^2\right] + \left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2$$

Bias and Variance

$$\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - f(\mathbf{x})\right)^2\right] = \mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)^2\right] + \left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2$$

$$\mathbb{E}_D[E_{out}(h^D)] = \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - f(\mathbf{x})\right)^2\right]\right] = \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)^2\right] + \left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2\right] =$$

$$= \mathbb{E}_{\mathbf{X}}[variance(\mathbf{x}) + bias(\mathbf{x})] = bias + variance$$

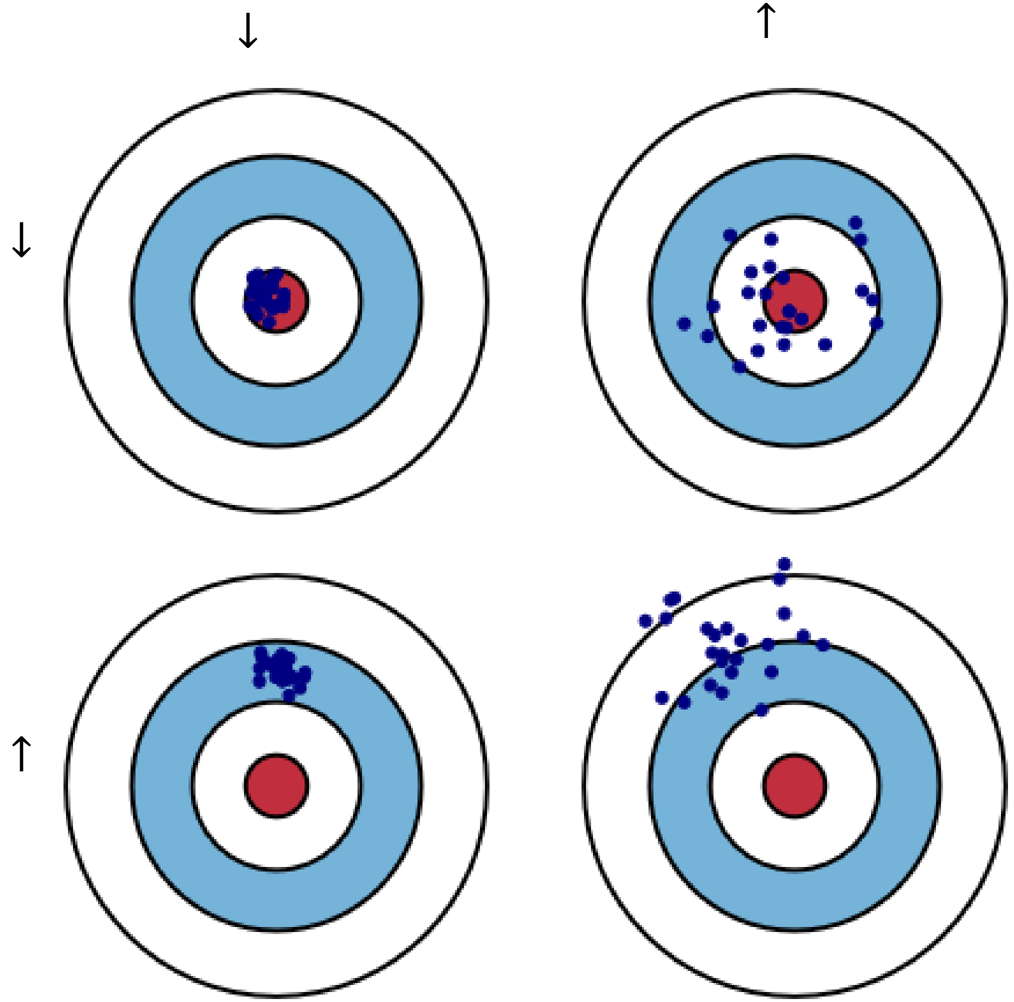
$$bias = \mathbb{E}_{\mathbf{X}}\left[\left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2\right]$$

$$variance = \mathbb{E}_{\mathbf{X}}\left[\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)^2\right]\right]$$

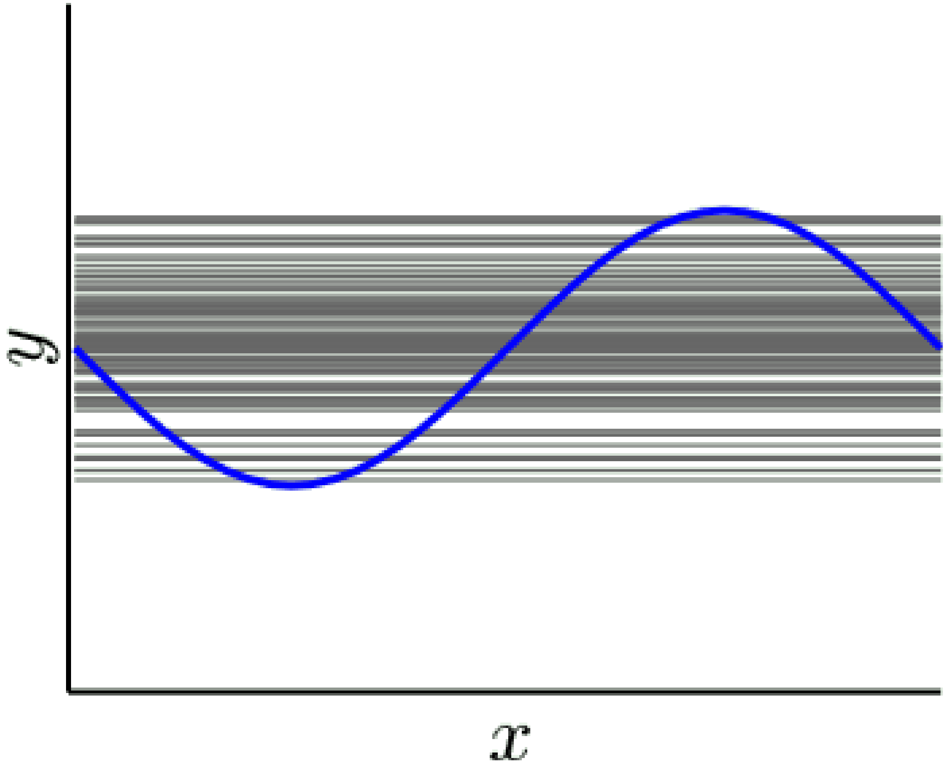
Bias and Variance

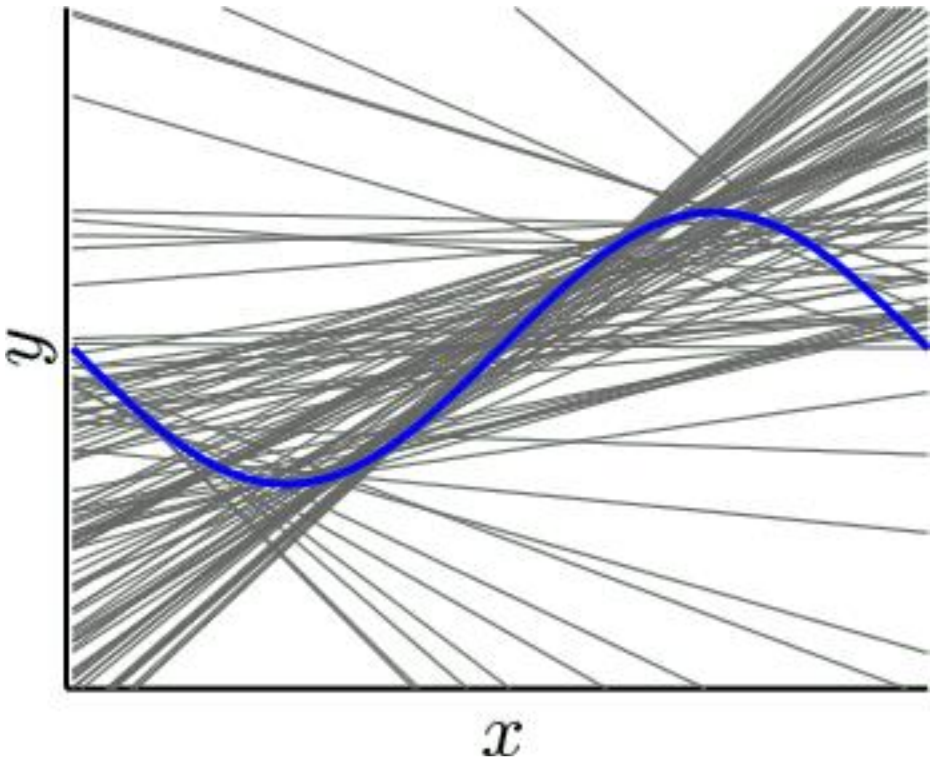
$$variance = \mathbb{E}_{\mathbf{x}}\left[\mathbb{E}_D\left[\left(h^D(\mathbf{x}) - \bar{h}(\mathbf{x})\right)^2\right]\right]$$

$$bias = \mathbb{E}_{\mathbf{x}}\left[\left(\bar{h}(\mathbf{x}) - f(\mathbf{x})\right)^2\right]$$

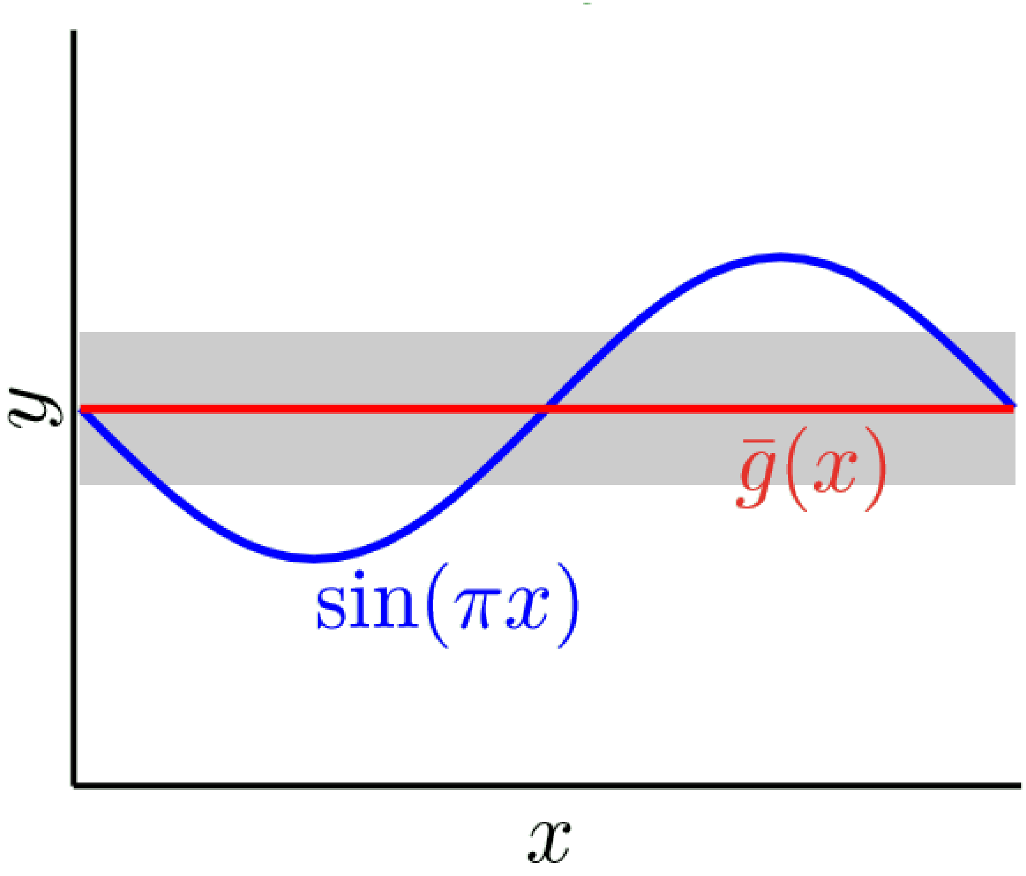


Цель - синус

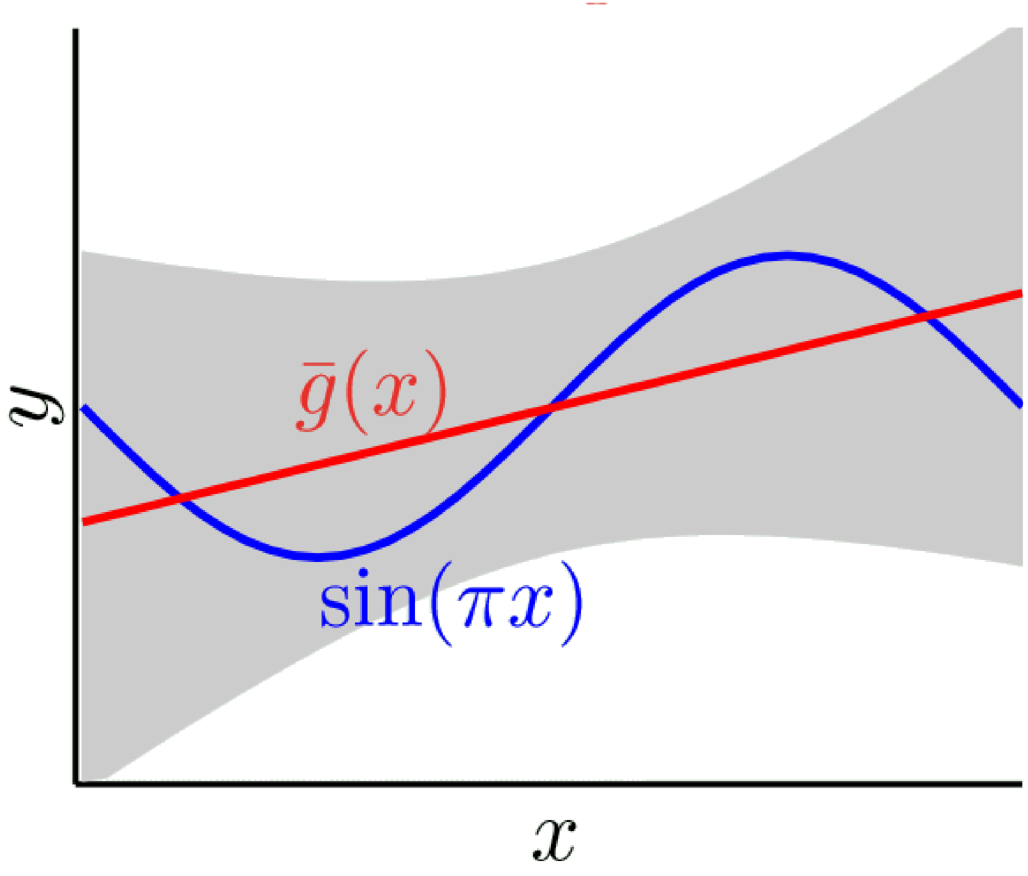




Цель - синус

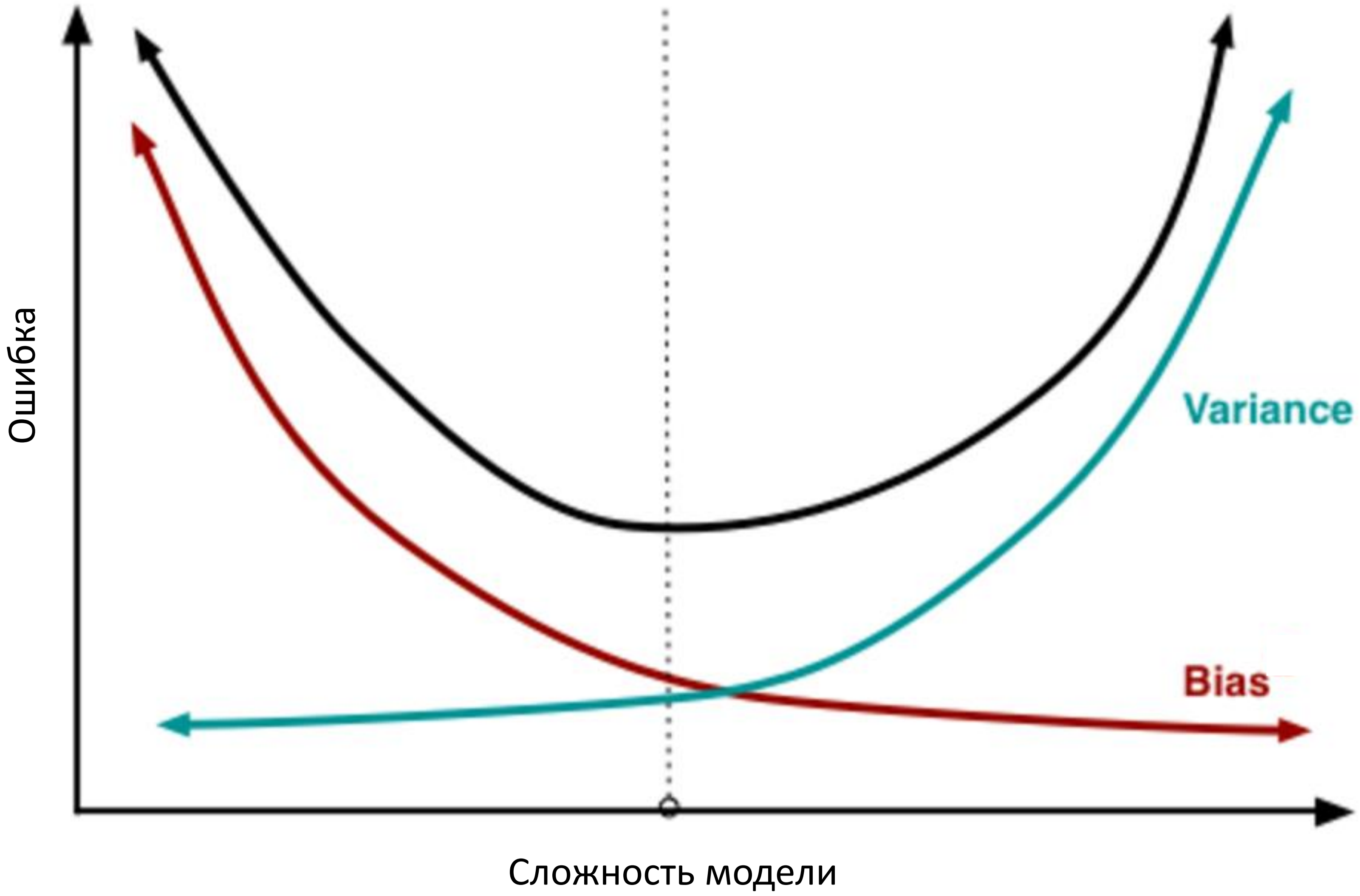


bias $= \mathbf{0.50}$ var $= \mathbf{0.25}$



bias $= \mathbf{0.21}$ var $= \mathbf{1.69}$

Bias and Variance



Регуляризация регрессии

Регрессия. Регуляризация.

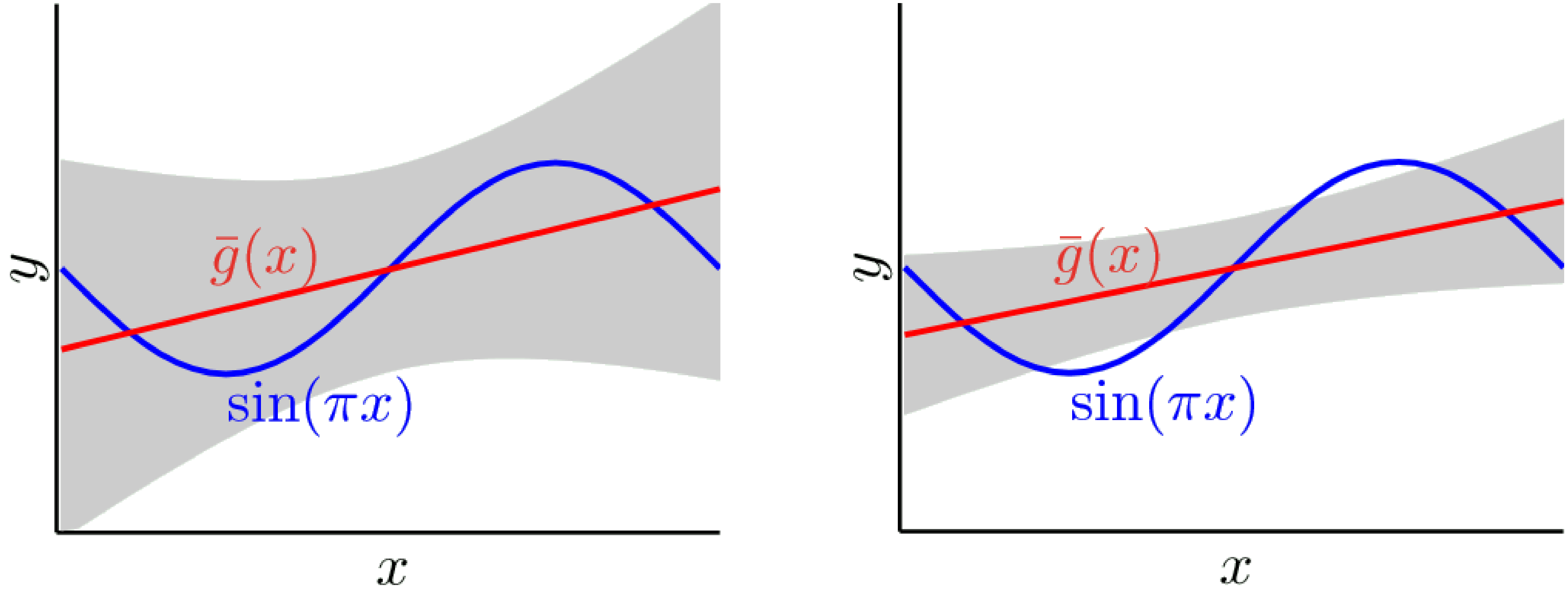
$$E_{in}(\mathbf{w}) = \frac{1}{N}\left|\left|\mathbf{Xw} - \mathbf{y}\right|\right|_2^2$$

$$E_{in}(\mathbf{w}) + \frac{\alpha}{\boldsymbol{N}}\mathbf{w}^{\mathrm{T}}\mathbf{w} = \frac{1}{N}\Big((\mathbf{Xw} - \mathbf{y})^{\mathrm{T}}(\mathbf{Xw} - \mathbf{y}) + \alpha\mathbf{w}^{\mathrm{T}}\mathbf{w}\Big)$$

$$\nabla E_{in}(\mathbf{w}) = \frac{2}{N}\Big(\mathbf{X}^{\mathrm{T}}(\mathbf{Xw} - \mathbf{y}) + \alpha\mathbf{w}\Big) = 0$$

$$\mathbf{w} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \alpha\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

Регуляризация и синус

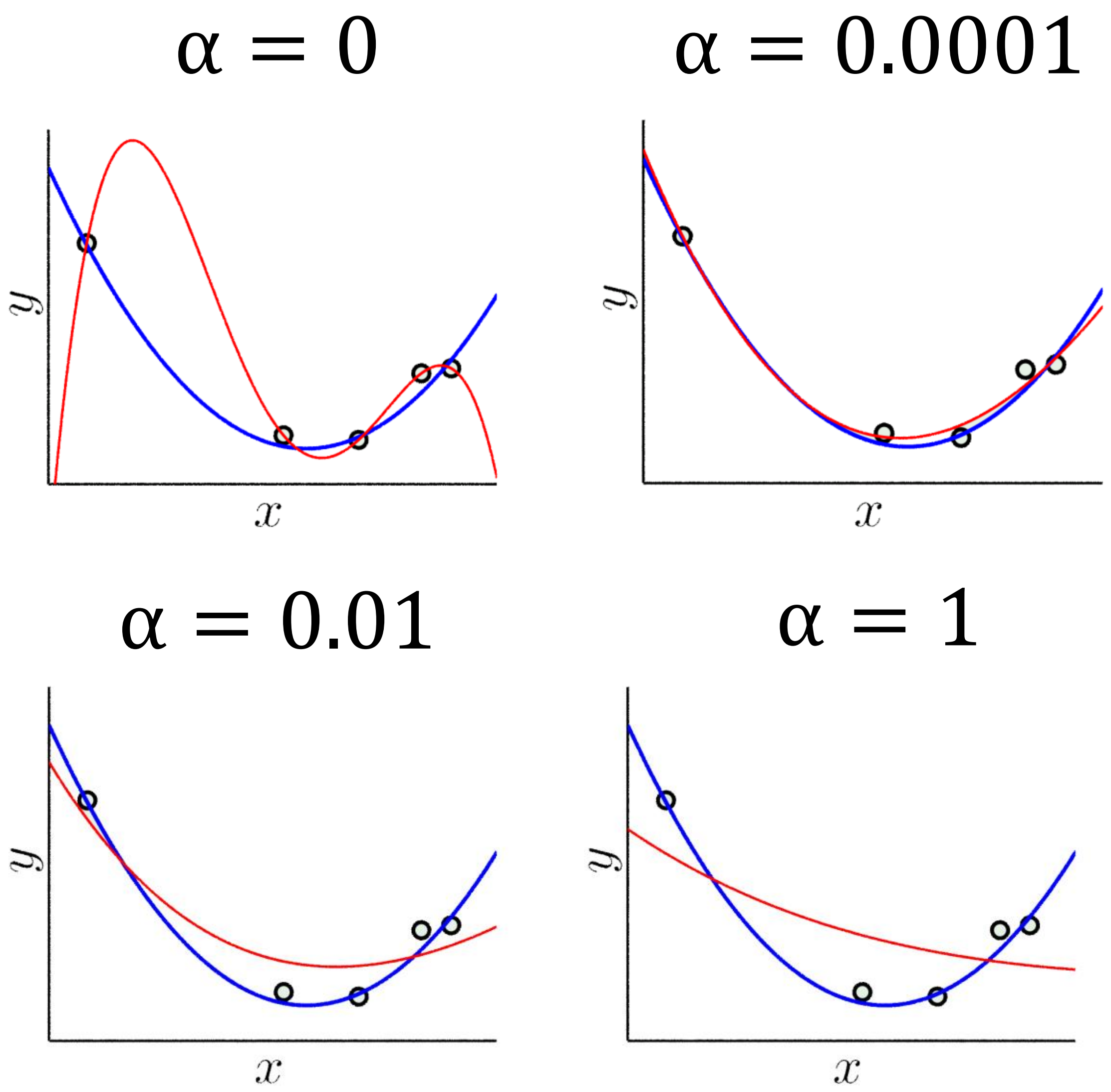


bias = **0.21** var = **1.69**

bias = **0.23** var = **0.33**

Переобучение и недообучение

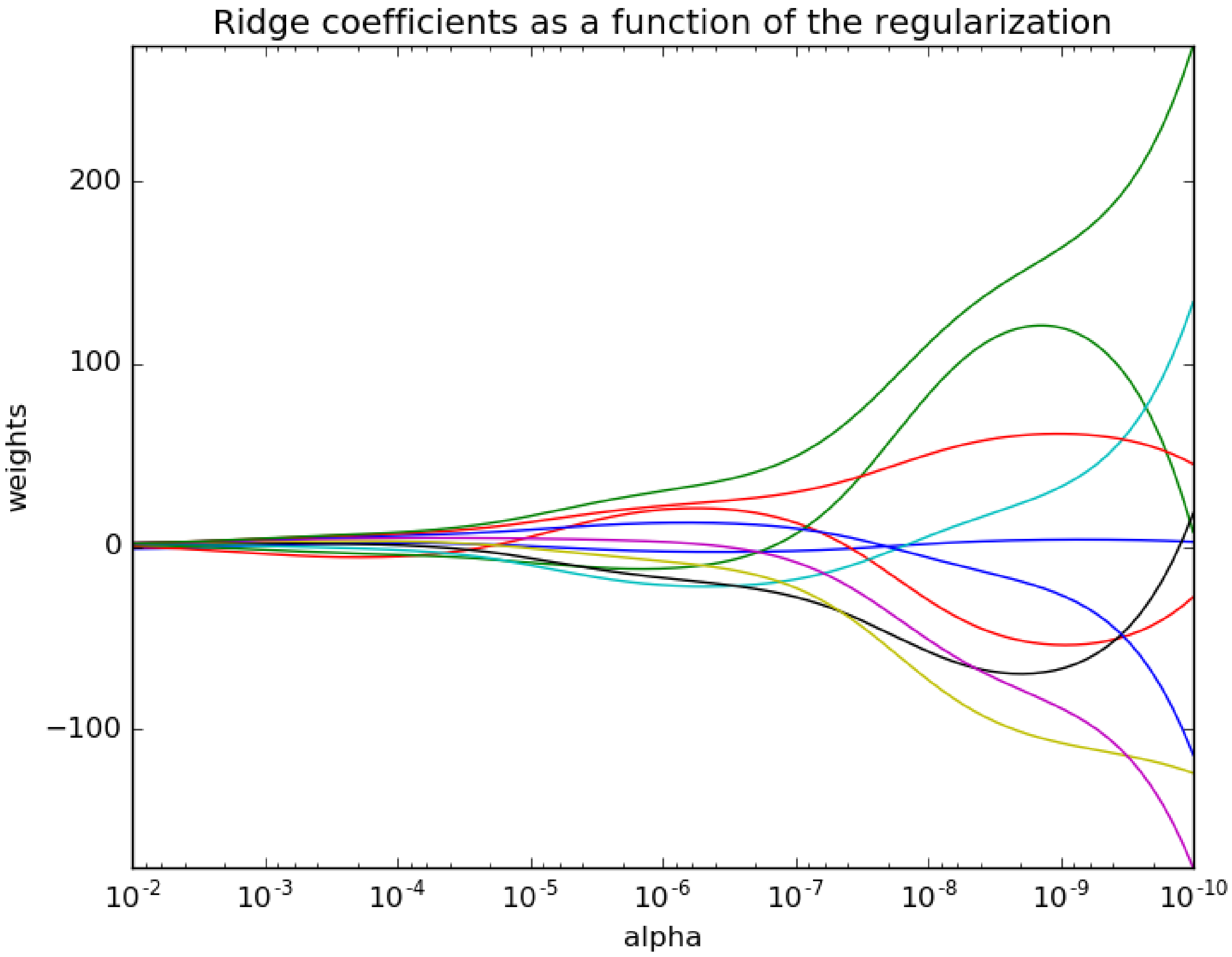
$$E_{in}(\mathbf{w}) + \frac{\alpha}{\boldsymbol{N}}\mathbf{w}^{\mathrm{T}}\mathbf{w}$$



Гребневая (Ridge) регрессия

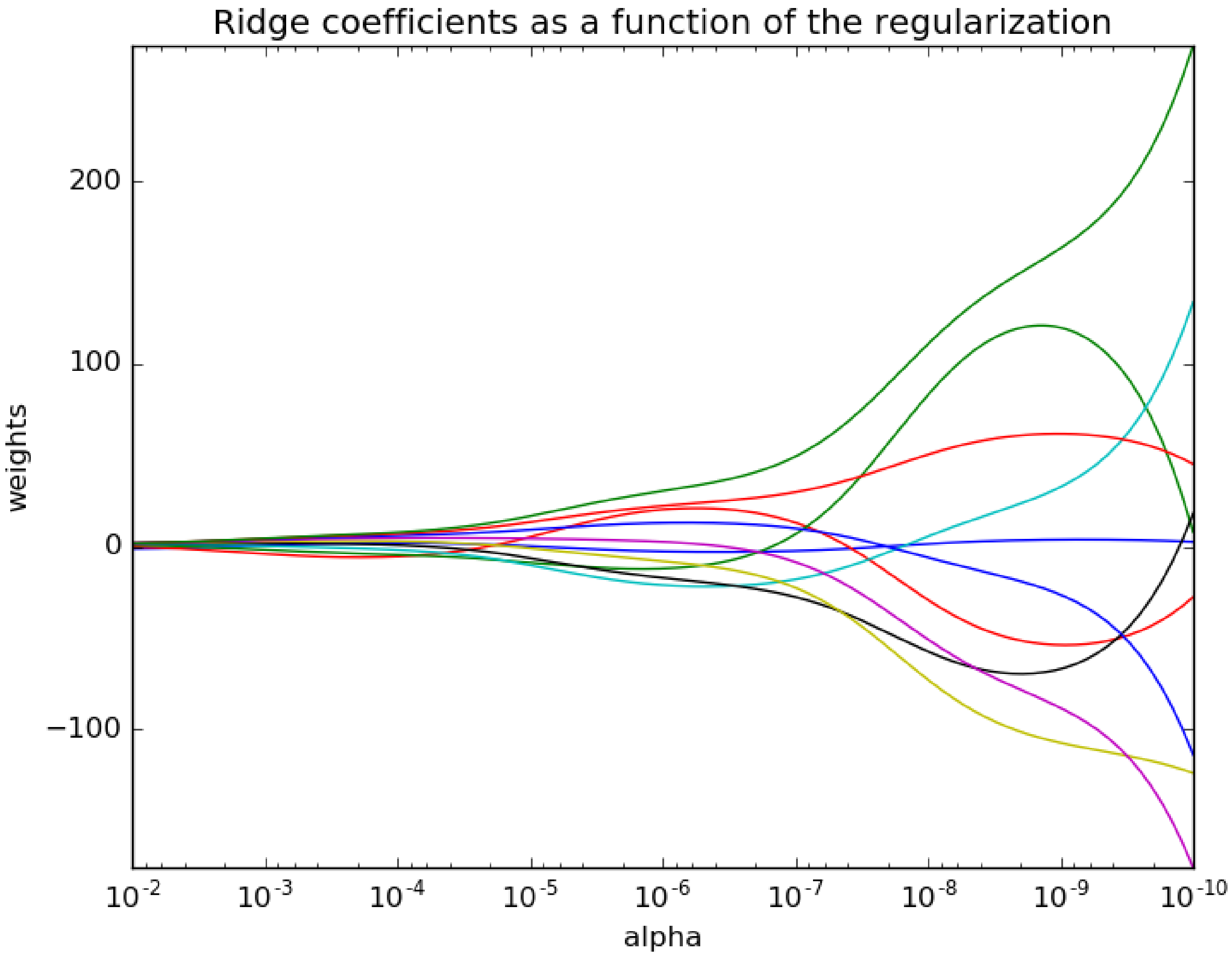
$$E_{in} = \frac{1}{N}\left|\left|\mathrm{X}\mathrm{w} - \mathrm{Y}\right|\right|_2^2 + \frac{\alpha}{N}\left|\left|\mathrm{w}\right|\right|_2^2$$

$$\mathrm{w} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X} + \alpha\mathrm{I}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathrm{Y}$$



Гребневая (Ridge) регрессия

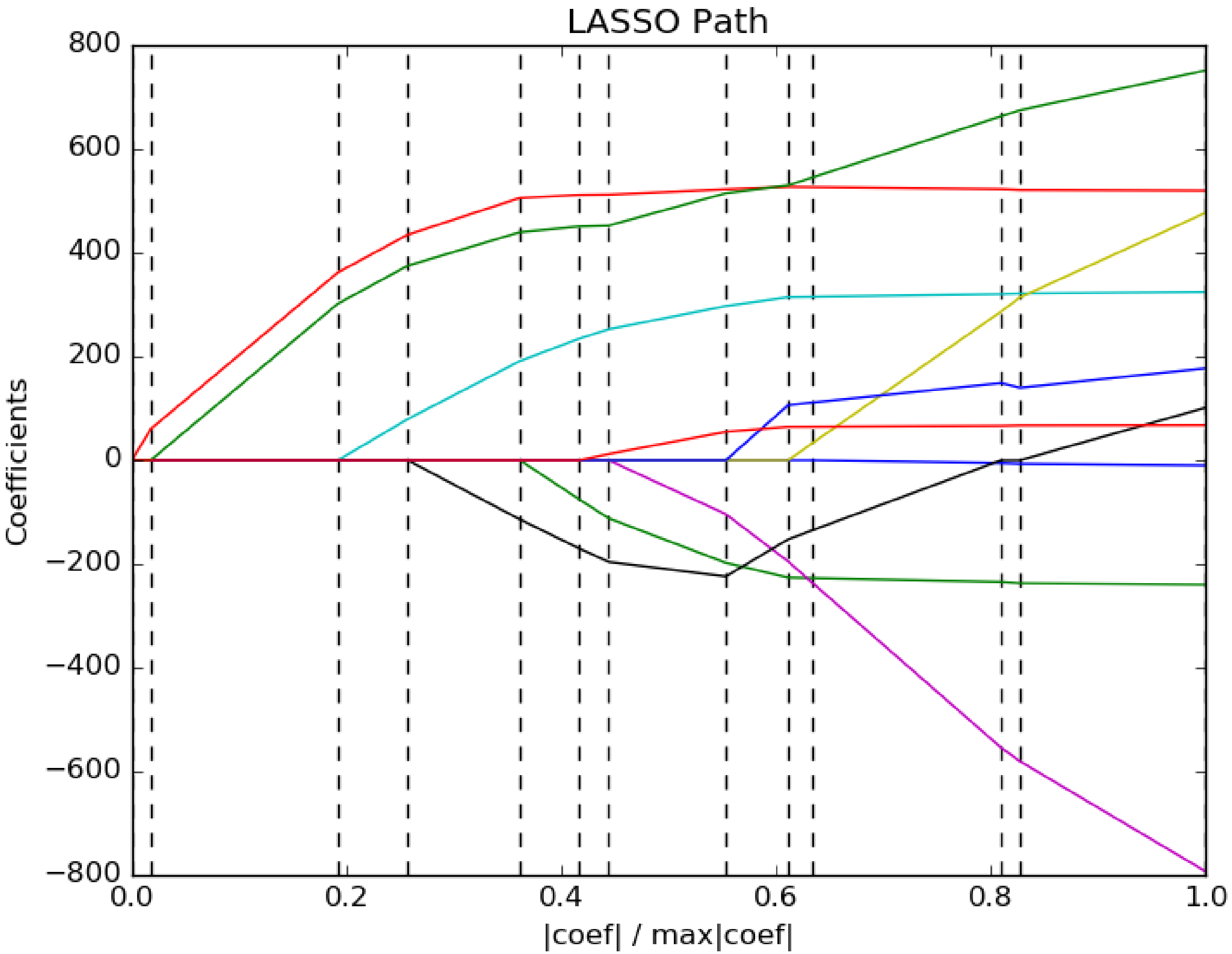
$$E_{in} = \frac{1}{2N}||Xw - Y||_2^2 + \alpha||w||_2^2$$



LASSO
(Least Absolute Shrinkage and Selection Operator)

$$E_{in} = \frac{1}{2N}||Xw - Y||_2^2 + \alpha||w||_1$$

Применяется покоординатный спуск или LARS.



LARS (Least Angle Regression)

1. Возьмем x_i наиболее сильно коррелирующий с y.

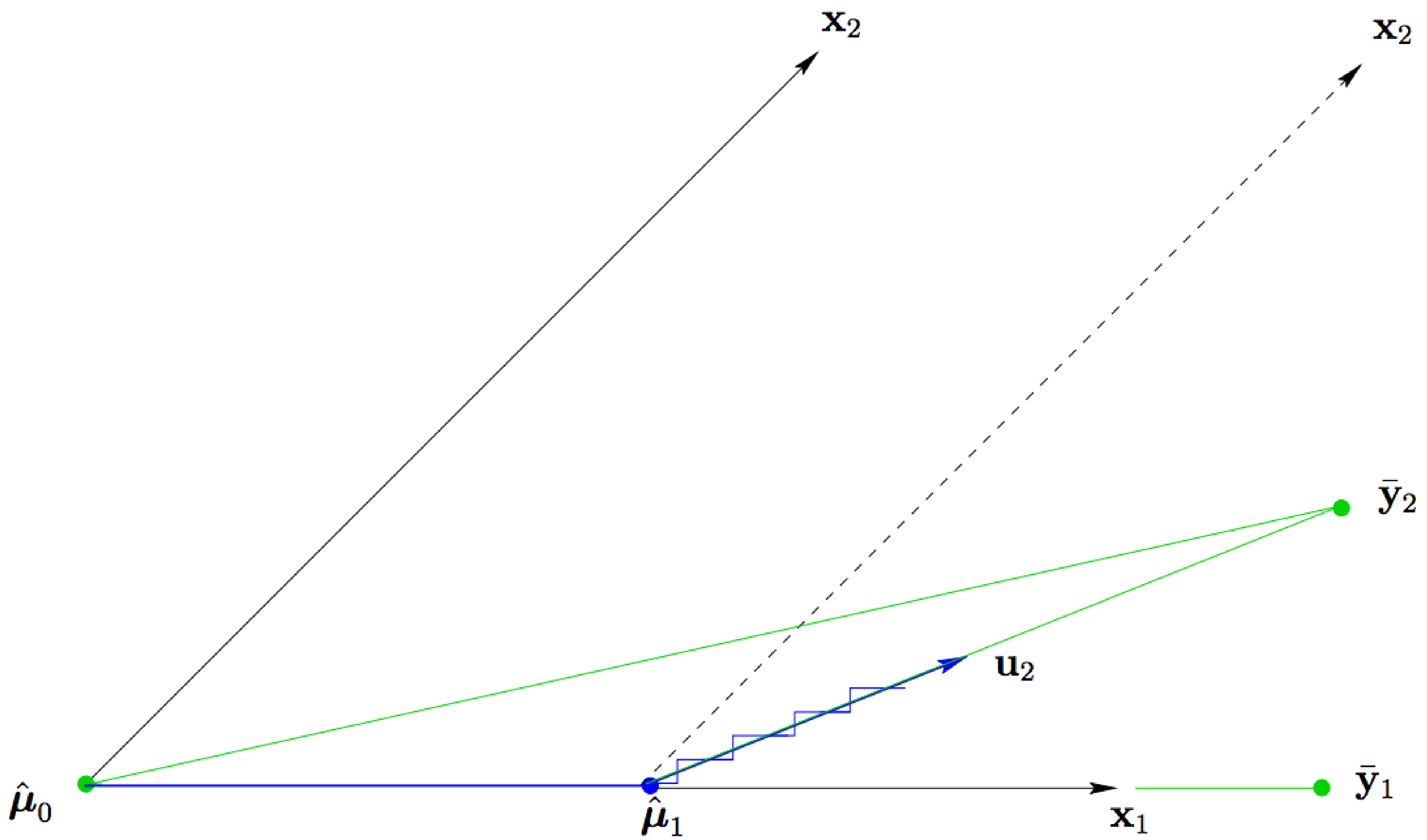
2. Сдвигаем коэффициент β_1 при x_i в сторону корреляции пока корреляция x_i с остаточным значением $r = y - \hat{y}$ остается максимальной.

3. В тот момент, как корреляция перестает быть максимальной, появляется еще один (или несколько) признак x_j с такой же корелляцией.

4. Вводим коэффициент β_2 при $\left(x_i \pm x_j\right)$.

5. $\rightarrow 2$

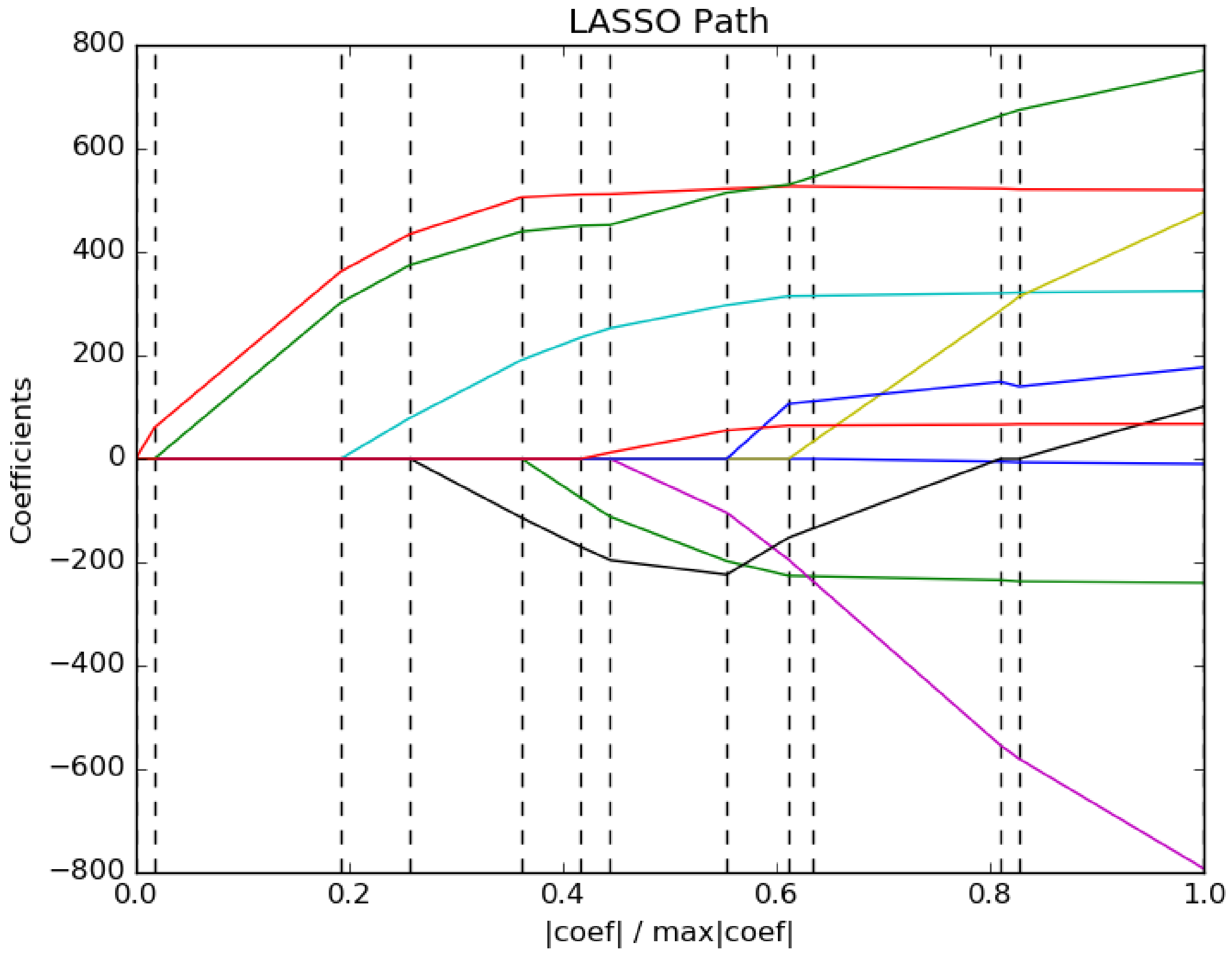
LARS


$\mathbf{x}_2$
$\mathbf{x}_2$
$\bar{\mathbf{y}}_2$
$\mathbf{u}_2$
$\hat{\boldsymbol{\mu}}_0$
$\hat{\boldsymbol{\mu}}_1$
$\mathbf{x}_1$
$\bar{\mathbf{y}}_1$

LASSO
(Least Absolute Shrinkage and Selection Operator)

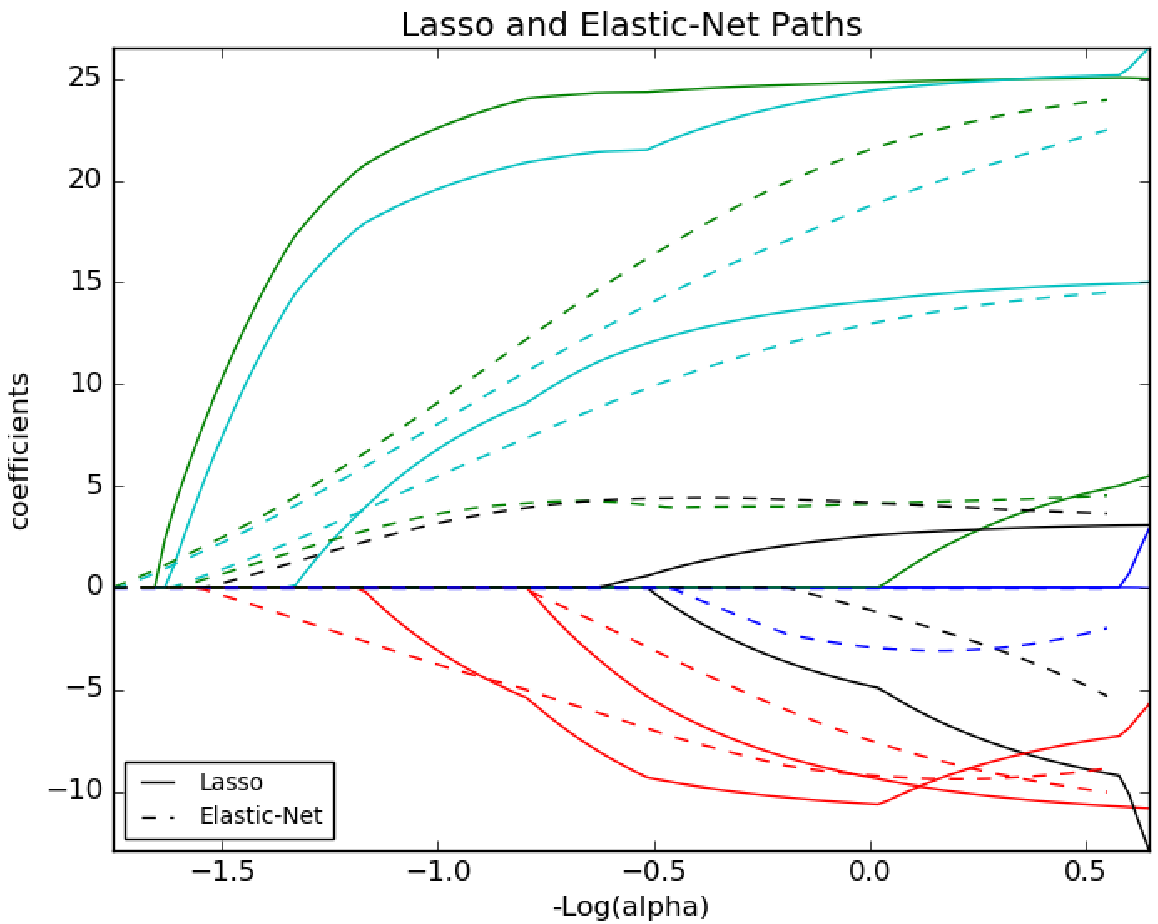
$$E_{in} = \frac{1}{2N}||Xw - Y||_2^2 + \alpha||w||_1$$

Применяется покоординатный спуск или LARS.



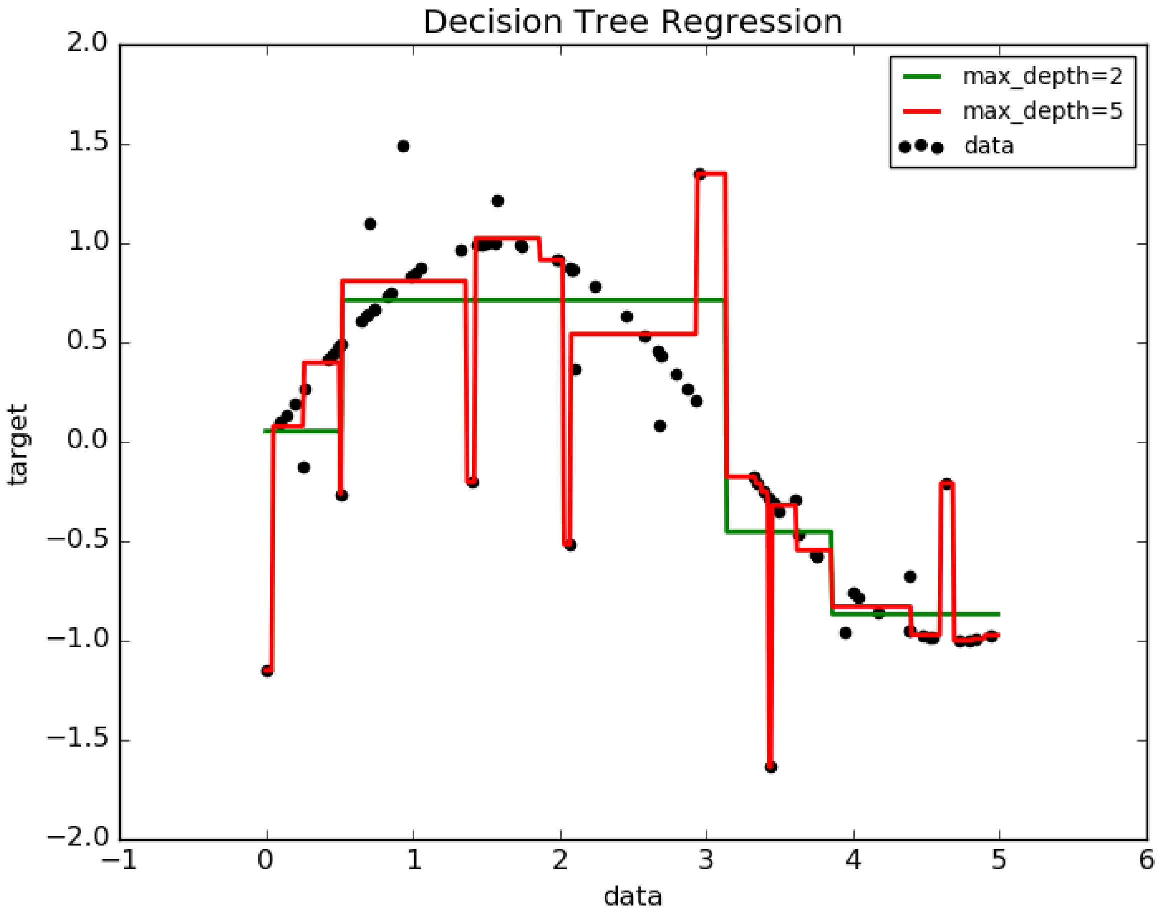
Elastic Net

$$E_{in} = \frac{1}{2N}\left|\left|\mathrm{Xw} - \mathrm{Y}\right|\right|_2^2 + \frac{1}{2}\alpha(1 - l1_{ratio})\left|\left|\mathrm{w}\right|\right|_2^2 + \alpha(l1_{ratio})\left|\left|\mathrm{w}\right|\right|_1$$



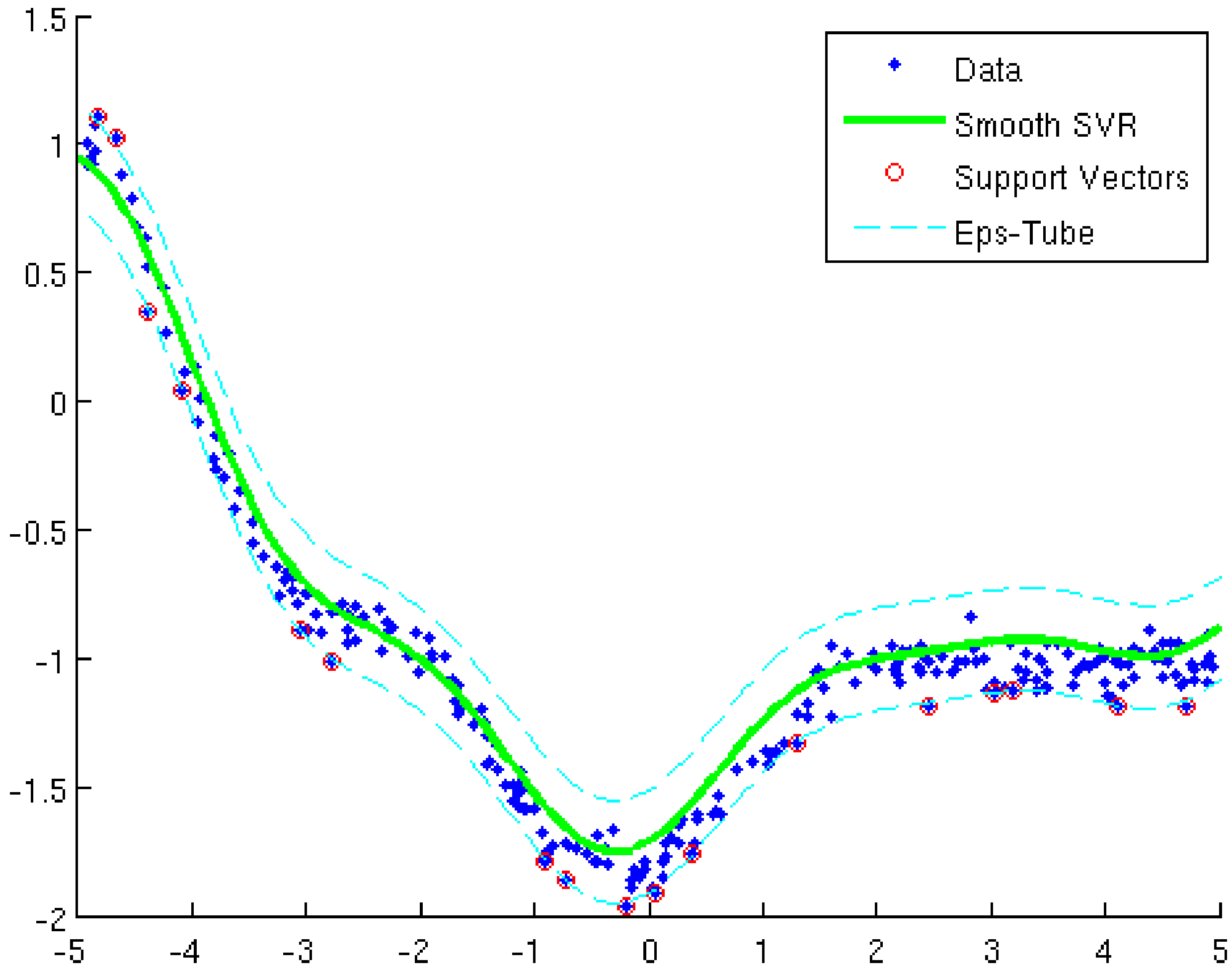
CART

$$\nabla V(\mathrm{R}) = \frac{1}{N^2}\sum\sum\frac{1}{2}\left(y_i - y_j\right)^2 - \left(\frac{1}{N_1^2}\sum\sum\frac{1}{2}\left(y'_i - y'_j\right)^2 + \frac{1}{N_2^2}\sum\sum\frac{1}{2}\left(y''_i - y''_j\right)^2\right)$$



Support Vector Regression Machine

$$\begin{cases} \frac{1}{2}\|\mathrm{w}\|^2 + C\sum(\xi_i + \xi_i^*) \to min \\ y_i - \mathrm{w}^{\mathrm{T}}\mathrm{x}_i - b \le \epsilon + \xi_i \\ \mathrm{w}^{\mathrm{T}}\mathrm{x}_i + b - y_i \le \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge \epsilon \end{cases}$$



R^2-score (коэффициент детерминации)

Абсолютное значение ошибки довольно неинформативно.

Удобная оценка результата регрессии:

$$R^2 = 1 - \frac{u}{v}$$

$$u = \sum (h(\mathrm{x}_i) - y_i)^2 \qquad v = \sum (\bar{y} - y_i)^2 \qquad \bar{y} = \frac{1}{N} \sum y_i$$