БАЗОВЫЕ ПРИНЦИПЫ МАШИННОГО ОБУЧЕНИЯ

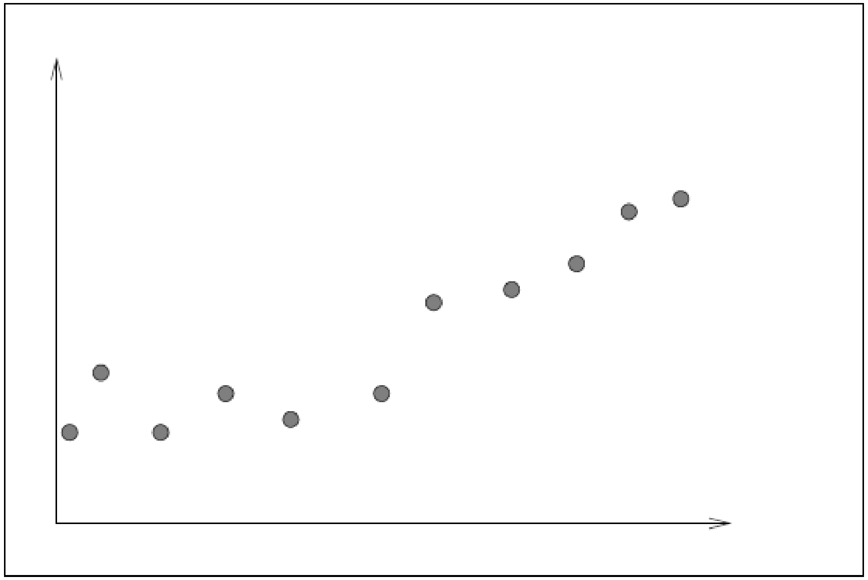
на примере линейной регрессии

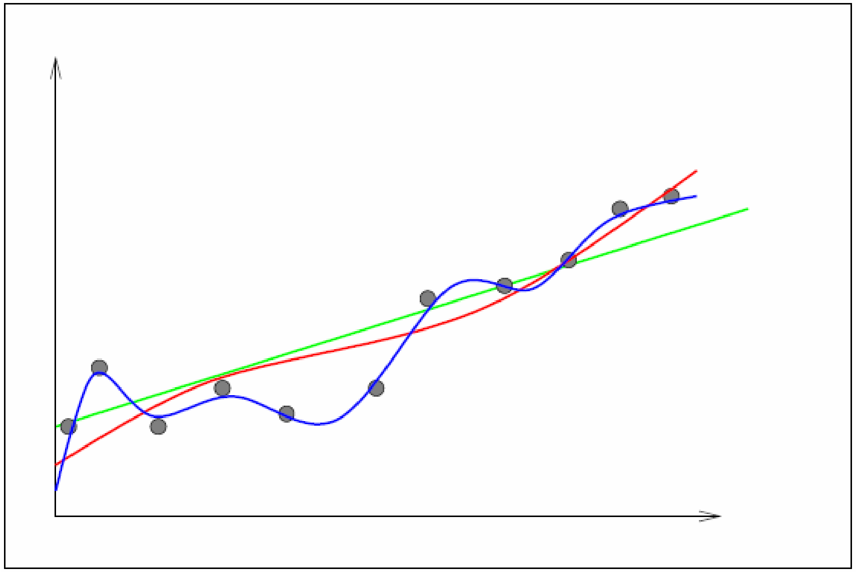
Регрессия

Дана обучающая выборка

$$X_N=\{(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_N,y_N)\},\qquad (\mathbf{x}_i,y_i)\in \mathbf{R}^P\times\mathbf{R}$$

Цель: для всех новых значений $\mathbf{x}$ оценить значения y





Метод наименьших квадратов

- Линейная модель: рассмотрим линейную функцию

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{p} x_j w_j = \mathbf{x}^\top \mathbf{w}, \quad \mathbf{x} = (1, x_1, \ldots, x_p).$$

- Таким образом, по вектору входов $\mathbf{x}^\top = (x_1, \ldots, x_p)$ мы будем предсказывать выход y как

$$\hat{y}(\mathbf{x}) = \hat{w}_0 + \sum_{j=1}^{p} x_j \hat{w}_j = \mathbf{x}^\top \hat{\mathbf{w}}.$$

Метод наименьших квадратов

- Как найти оптимальные параметры $\hat{\mathbf{w}}$ по тренировочным данным вида $(\mathbf{x}_i, y_i)_{i=1}^{N}$?
- Метод наименьших квадратов: будем минимизировать

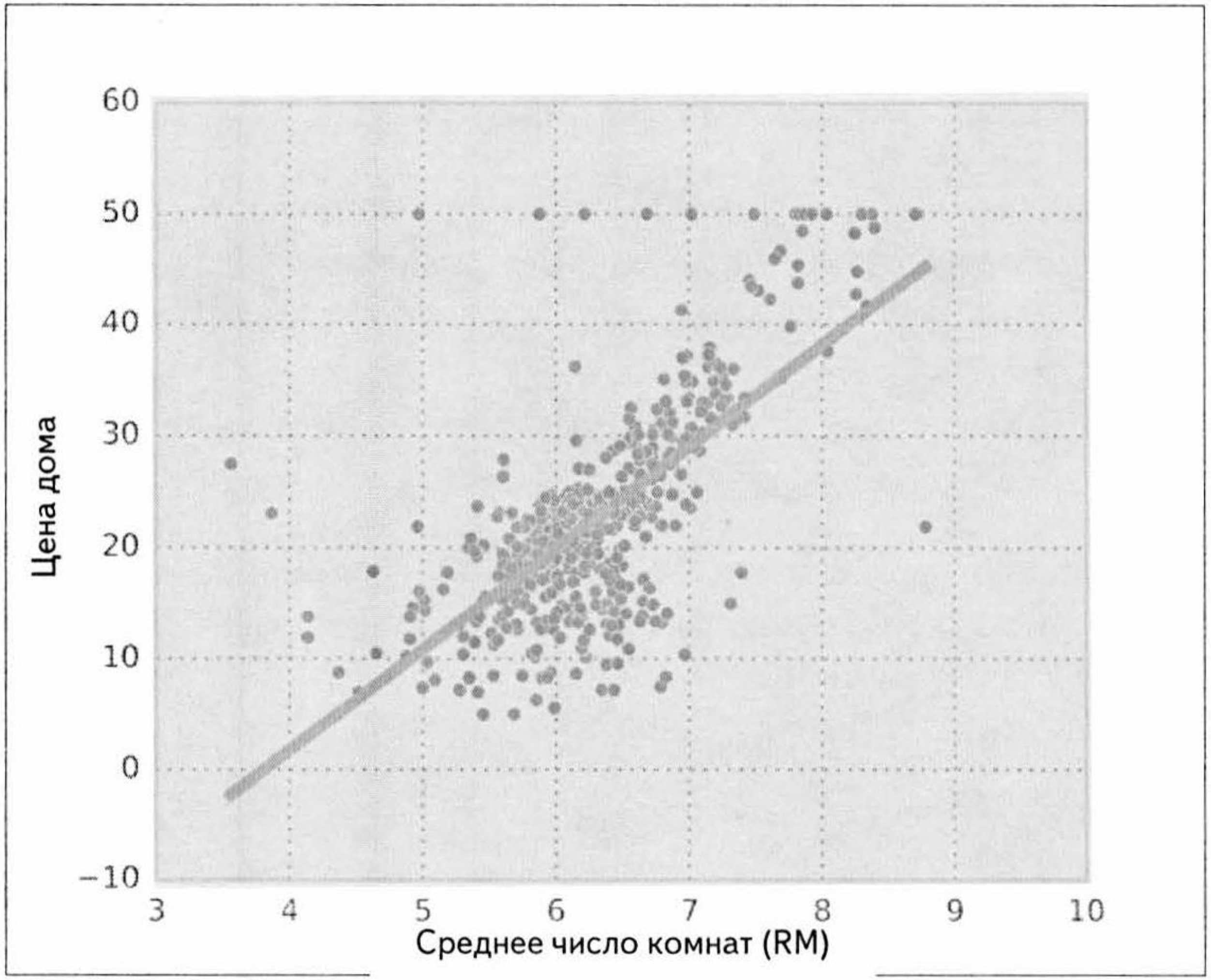
$$\mathrm{RSS}(\mathbf{w}) = \sum_{i=1}^{N} (y_i - \mathbf{x}_i^\top \mathbf{w})^2.$$

$$\mathrm{RSS}(\mathbf{w}) = (\mathbf{y} - \mathbf{X}\mathbf{w})^\top (\mathbf{y} - \mathbf{X}\mathbf{w}),$$

$$\hat{\mathbf{w}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y},$$

если матрица $\mathbf{X}^\top \mathbf{X}$ невырожденная.

Пример: прогнозирование стоимости домов



RMSE=6.6
R^2=0.31

Измерение ошибки в задачах регрессии

$$L(y, \hat{y}) = (y - \hat{y})^2$$

$$L(y, \hat{y}) = |y - \hat{y}|$$

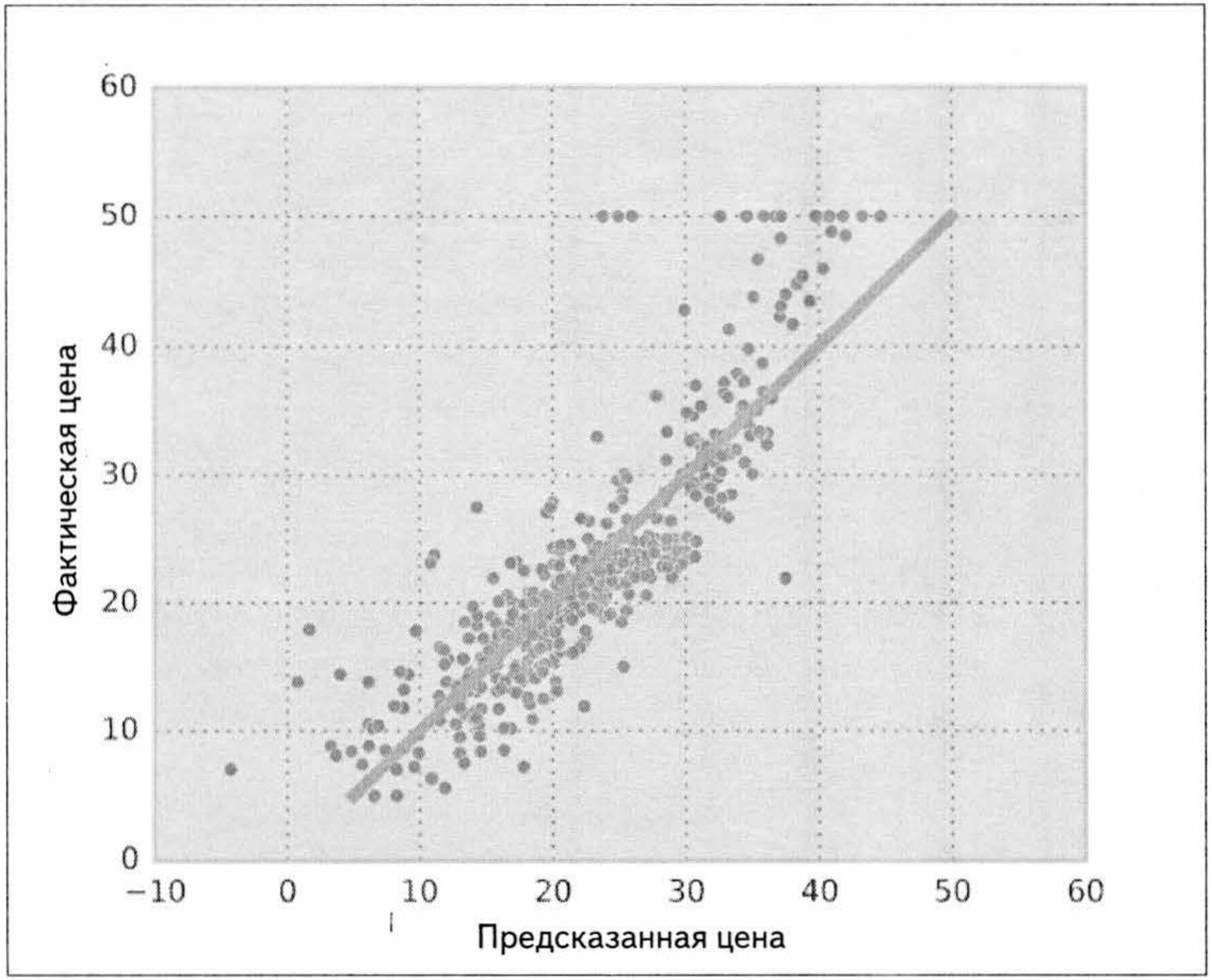
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}\ \ )^2}$$

Многомерная регрессия



RMSE=4.7
R^2=0.74

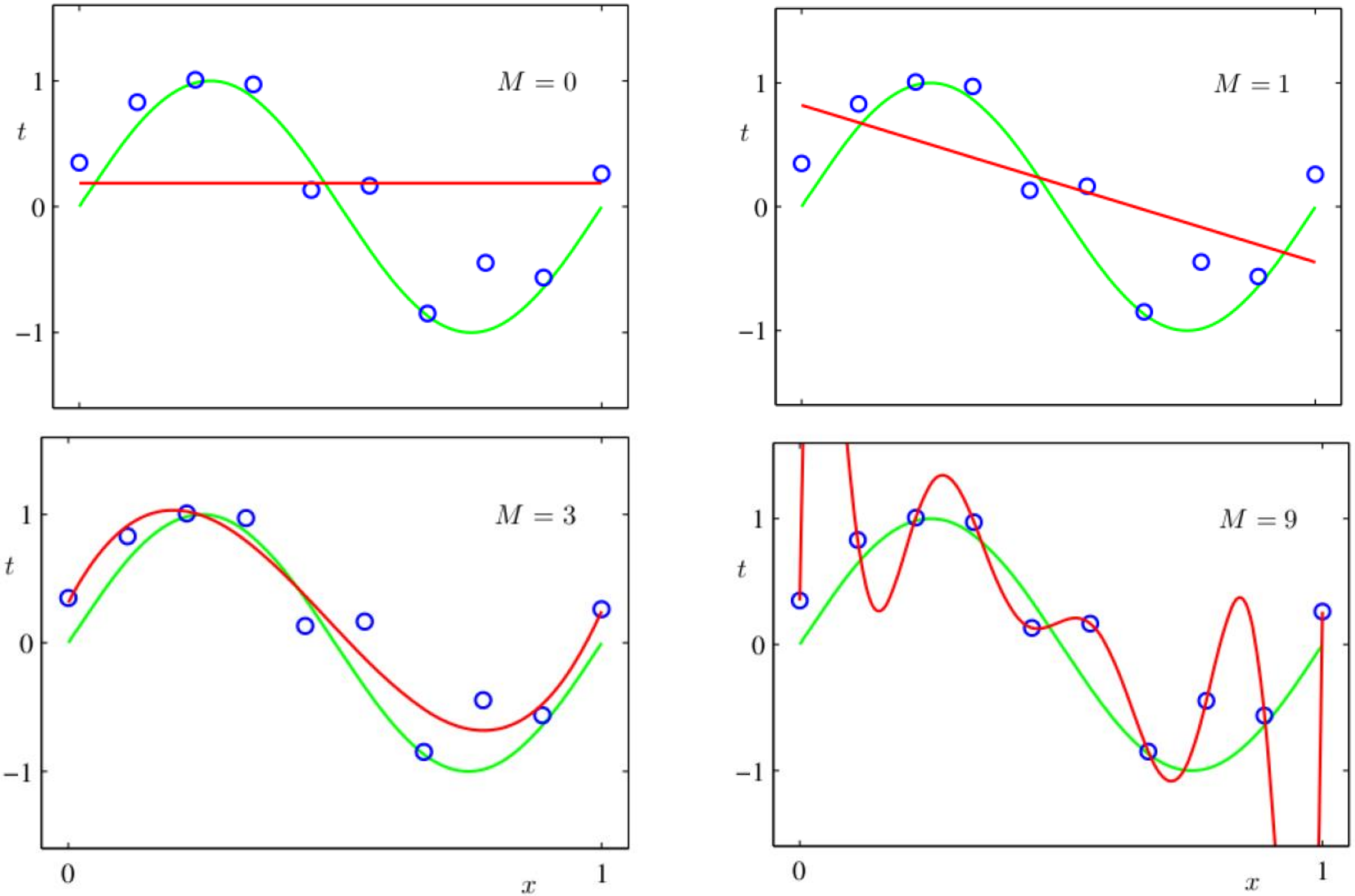
Регрессия, линейная по параметрам

$$f(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M} w_j \phi_j(\mathbf{x}) = \mathbf{w}^\top \boldsymbol{\phi}(\mathbf{x}).$$

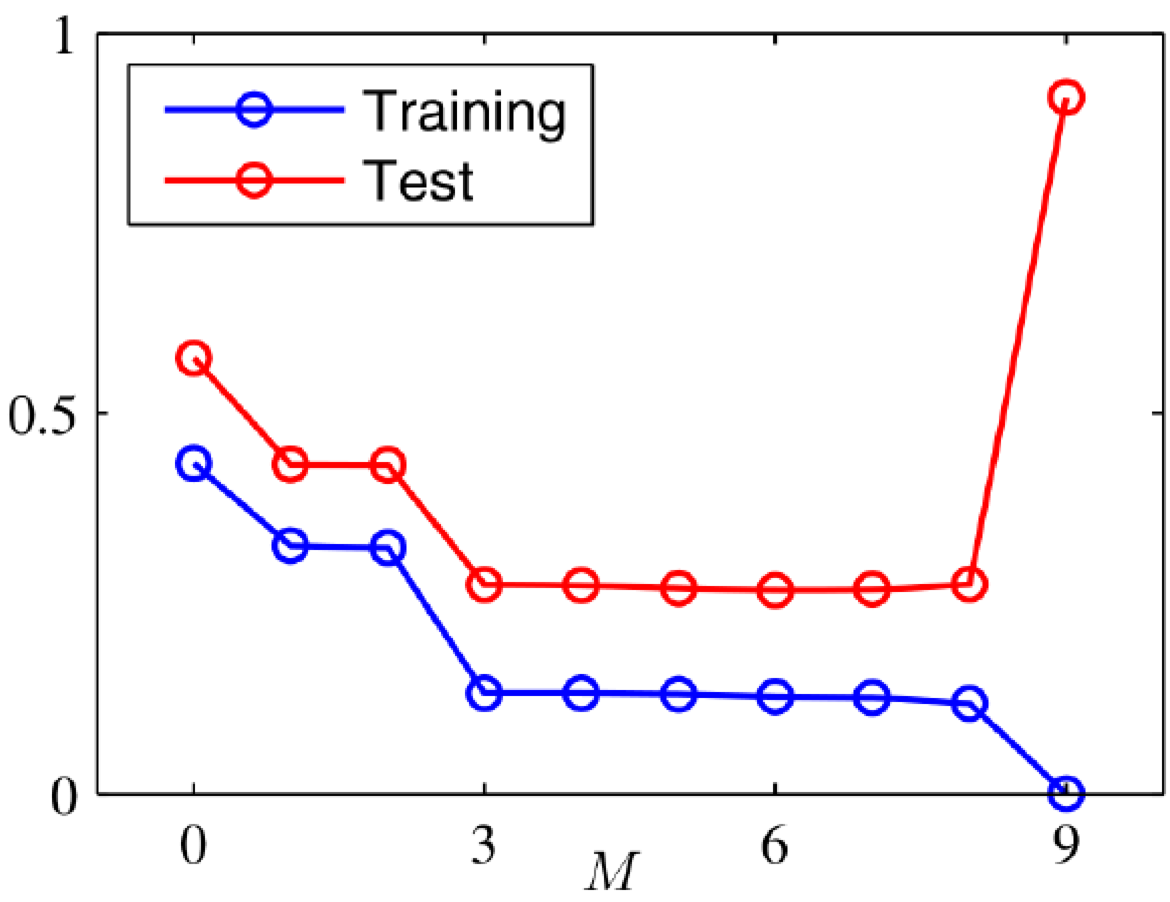
Например:

$$f(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M.$$

Полиномиальная регрессия



Значения $RMSE$



Значения коэффициентов

	$M = 0$	$M = 1$	$M = 6$	$M = 9$
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

L2- Регуляризация (гребневая регрессия)

- Было (для тестовых примеров $\{(x_i, y_i)\}_{i=1}^N$):

$$\mathrm{RSS}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, \mathbf{w}) - y_i)^2$$

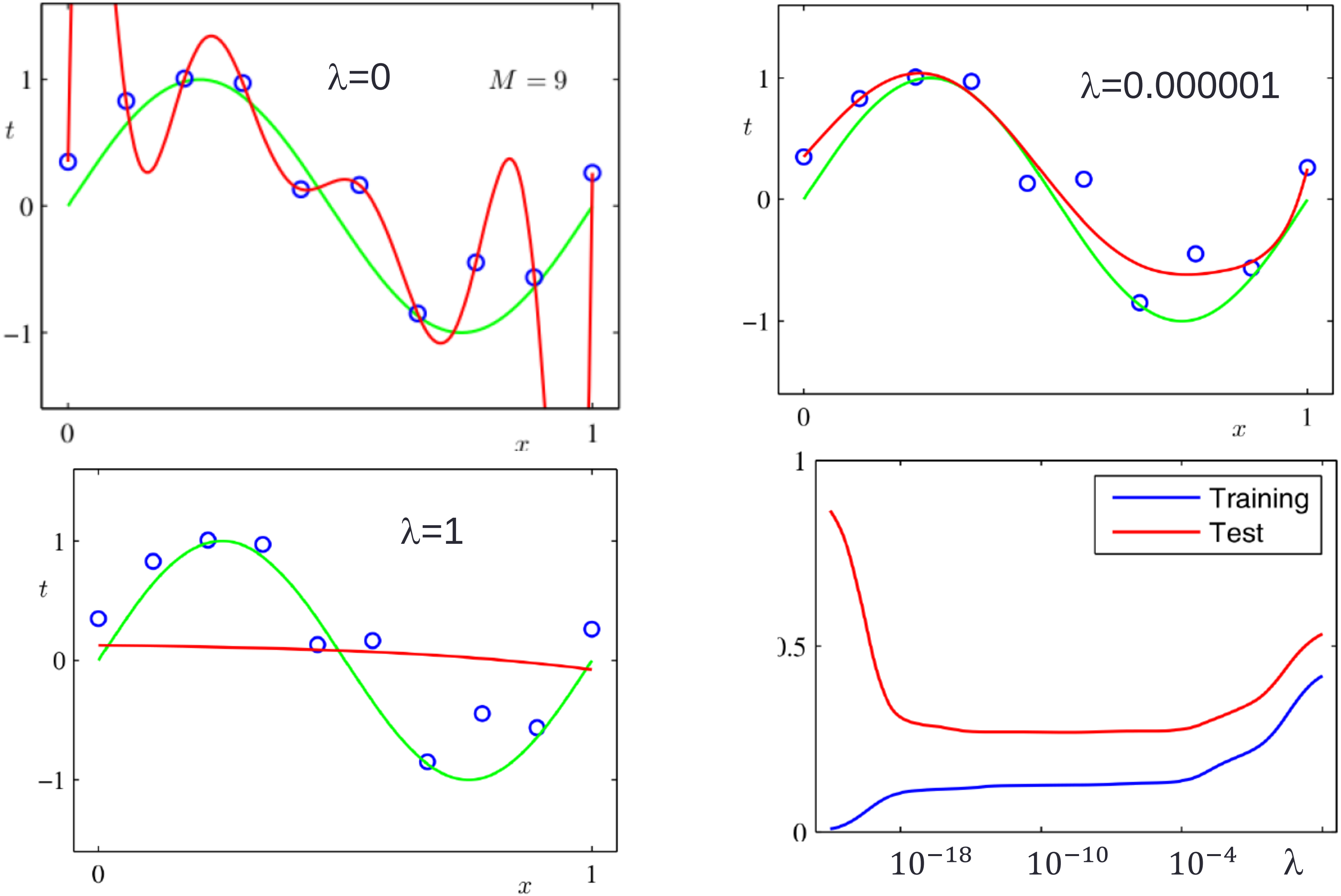
- Стало:

$$\mathrm{RSS}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, \mathbf{w}) - y_i)^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2,$$

где λ – коэффициент регуляризации

В регрессии, линейной по факторам: $$\mathbf{w}^* = \left(\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I}\right)^{-1} \mathbf{X}^\top \mathbf{y}.$$

L2- регуляризация



Коэффициенты гребневой регрессии

	λ=0	λ= 0.000001	λ=1
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01

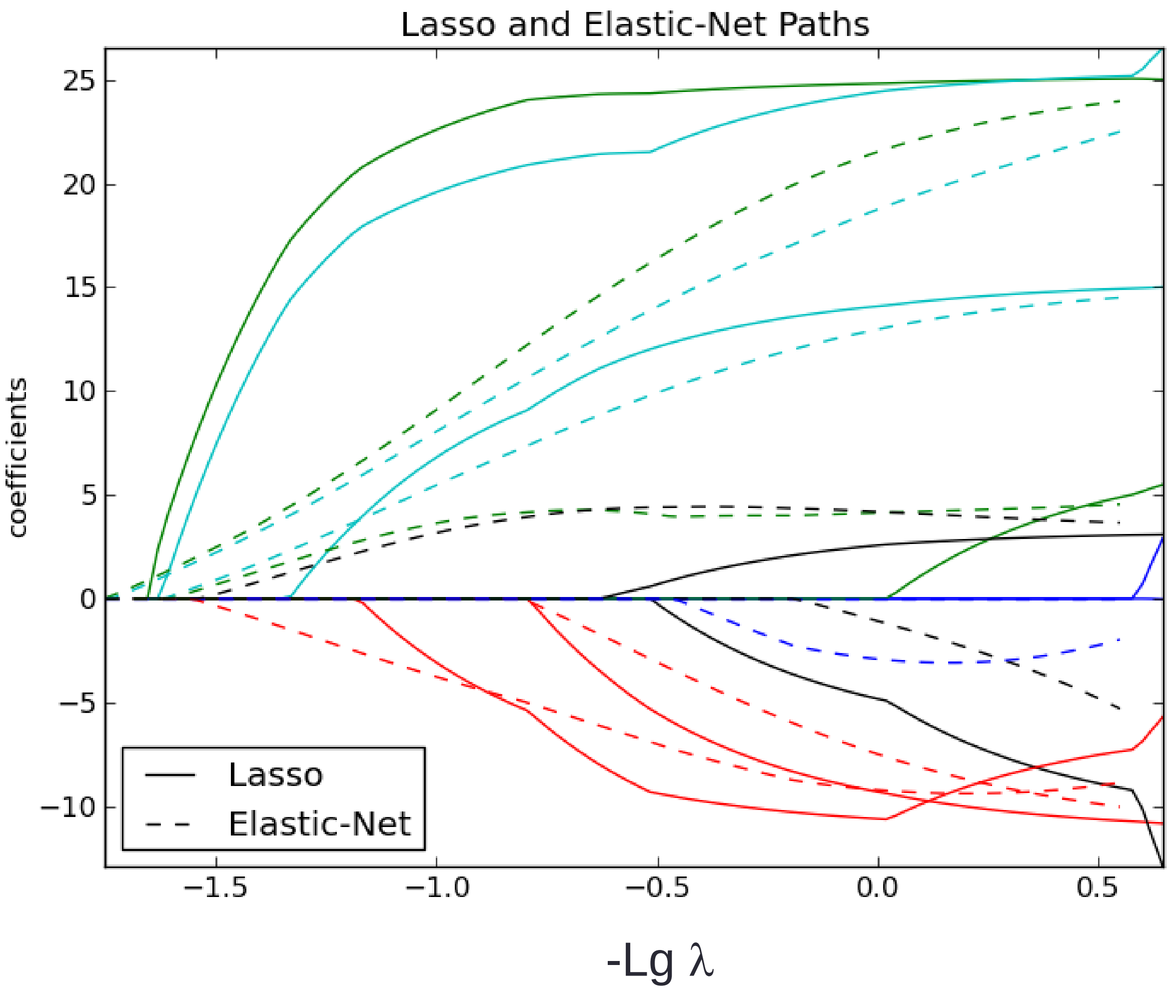
L1- регуляризация (Lasso)

$$\mathrm{RSS}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, \mathbf{w}) - y_i)^2 + \lambda\sum_{j=0}^{M}|w_j|.$$

Эластичная сеть:

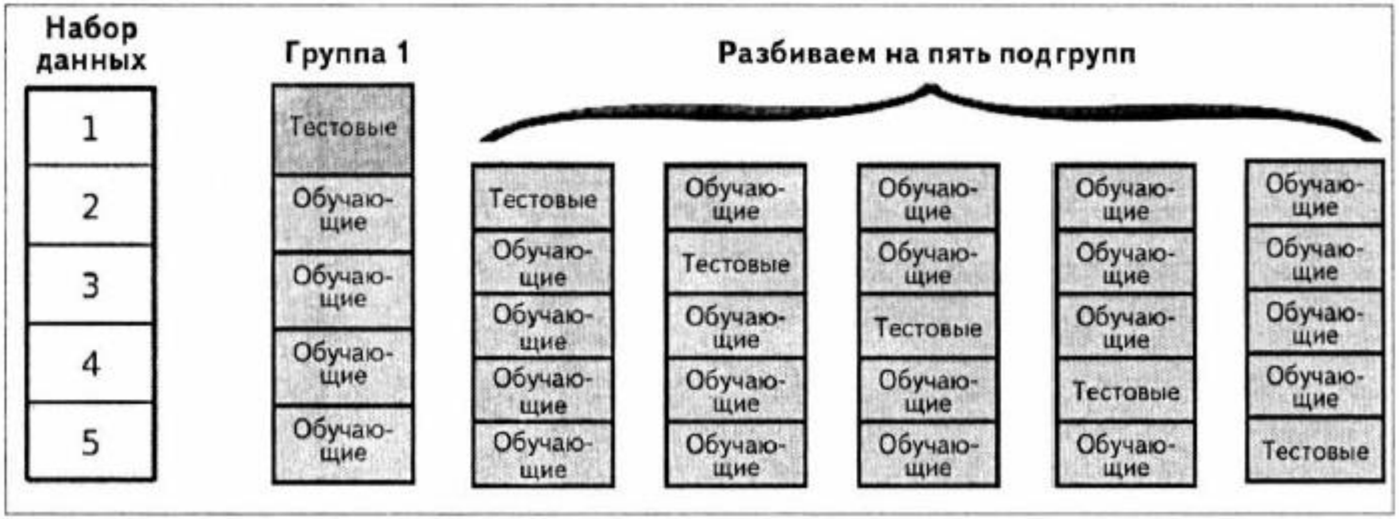
$$\mathrm{RSS}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(x_i, \mathbf{w}) - y_i)^2 + \lambda_1\sum_{j=0}^{M}|w_j| + \frac{\lambda_2}{2}\|\mathbf{w}\|^2$$

Пример использования Lasso и ElasticNet



Настройка гиперпараметров

- 1) три выборки – обучающая (настраиваются параметры); валидационная (настраиваются гиперпараметры) и тестовая (анализируются результаты обучения)
- 2) кросс-валидация (перекрестная проверка):



Спасибо за внимание!