Машинное обучение
Методы восстановления регрессии

Смирнов Илья

Руководитель направления

Содержание

- Метод наименьших квадратов
- Геометрический смысл
- Регуляризация
- Сингулярное разложение

Метод наименьших квадратов

X = $\mathbf{R}^n$, Y = $\mathbf{R}$

Модель: $\mathrm{a}(\mathrm{x}) = f(x, \alpha)$

Метод наименьших квадратов (МНК):

$$Q(\alpha, X^l) = \sum_{i=1}^{l} w_i (f(x_i, \alpha) - y_i)^2 \rightarrow min \text{ по } \alpha$$

w_i — вес, степень важности i-го объекта

Многомерная линейная регрессия

$f_1(x), \dots, f_n(x)$ — числовые признаки;

Модель $$f(x,\alpha) = \sum_{j=1}^{n} \alpha_j f_j(x), \qquad \alpha \in \mathbb{R}^n$$

Матричная форма:

$$\underset{\ell\times n}{F} = \begin{pmatrix} f_1(x_1) & \dots & f_n(x_1) \\ \dots & \dots & \dots \\ f_1(x_\ell) & \dots & f_n(x_\ell) \end{pmatrix}, \quad \underset{\ell\times 1}{y} = \begin{pmatrix} y_1 \\ \dots \\ y_\ell \end{pmatrix}, \quad \underset{n\times 1}{\alpha} = \begin{pmatrix} \alpha_1 \\ \dots \\ \alpha_n \end{pmatrix}$$

$$Q(\alpha, X^\ell) = \sum_{i=1}^{\ell} \big(f(x_i,\alpha) - y_i\big)^2 = \|F\alpha - y\|^2 \to \min_{\alpha}$$

Система уравнений

Необходимое условие минимума:

$$\frac{\partial Q}{\partial \alpha}(\alpha) = 2F^{\mathsf{T}}(F\alpha - y) = 0$$

$$F^{\mathsf{T}}F\alpha = F^{\mathsf{T}}y$$

$F^{T}F$ — ковариационная матрица n×n набора признаков $f_1,\ldots,f_n$

Решение системы: $\alpha^* = (F^{\mathsf{T}}F)^{-1}F^{\mathsf{T}}y = F^{+}y$

Значение функционала:

$$Q(\alpha^*) = \|P_F y - y\|^2$$

где P_F - проекционная матрица:

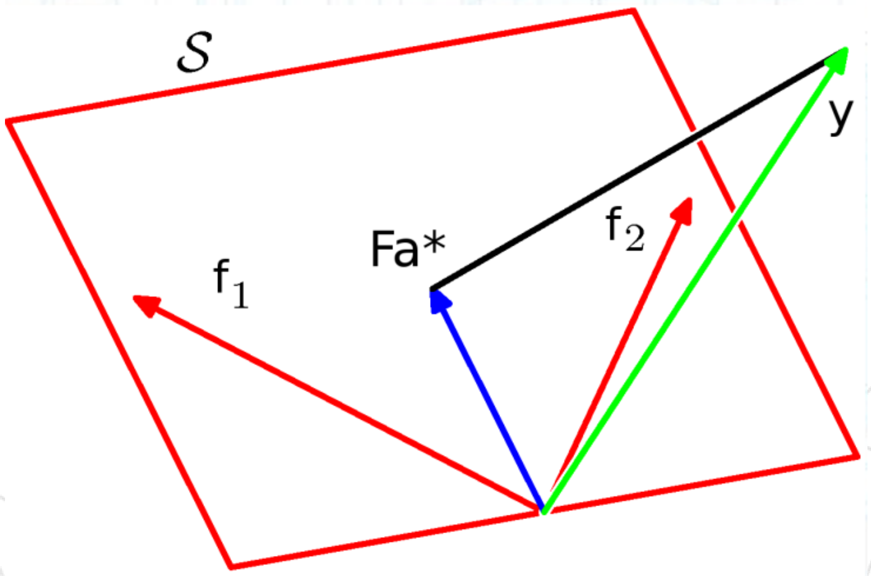
$$P_F = FF^{+} = F(F^{\mathsf{T}}F)^{-1}F^{\mathsf{T}}$$

Геометрический смысл

- Любой вектор вида **y = Fα** – линейная комбинация признаков

$$\|F\alpha - y\|^2 \to \min_{\alpha}$$

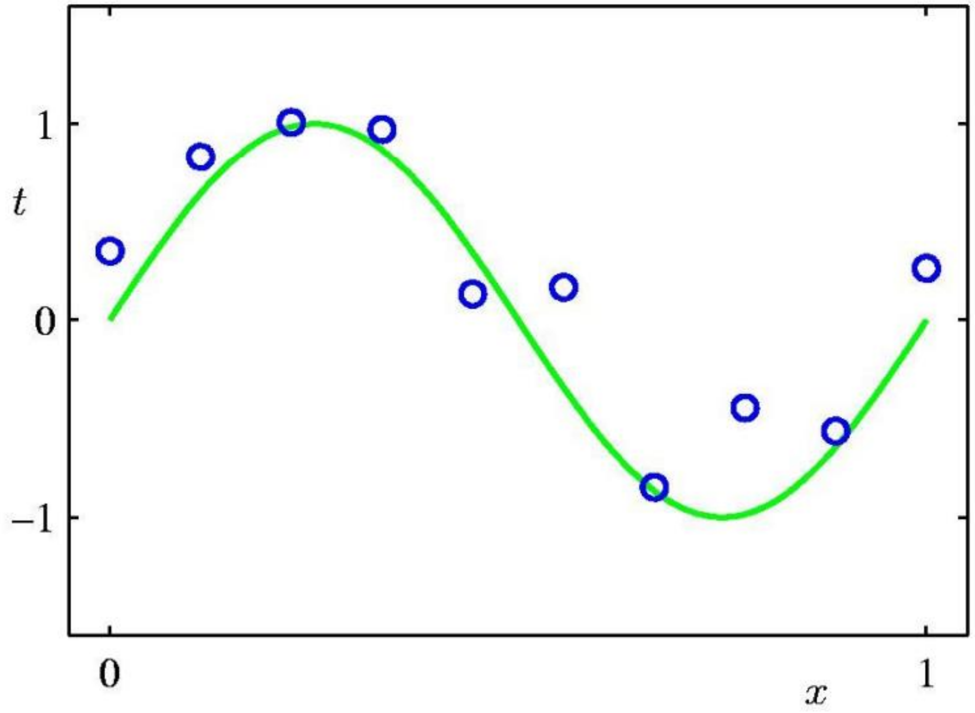
- **F**$\boldsymbol{\alpha}^*$ = аппроксимация вектора **y** с наименьшим квадратом тогда и только тогда, когда **F**$\boldsymbol{\alpha}^*$ - проекция вектора **y** на подпространство признаков



Пример: аппроксимация многочленами

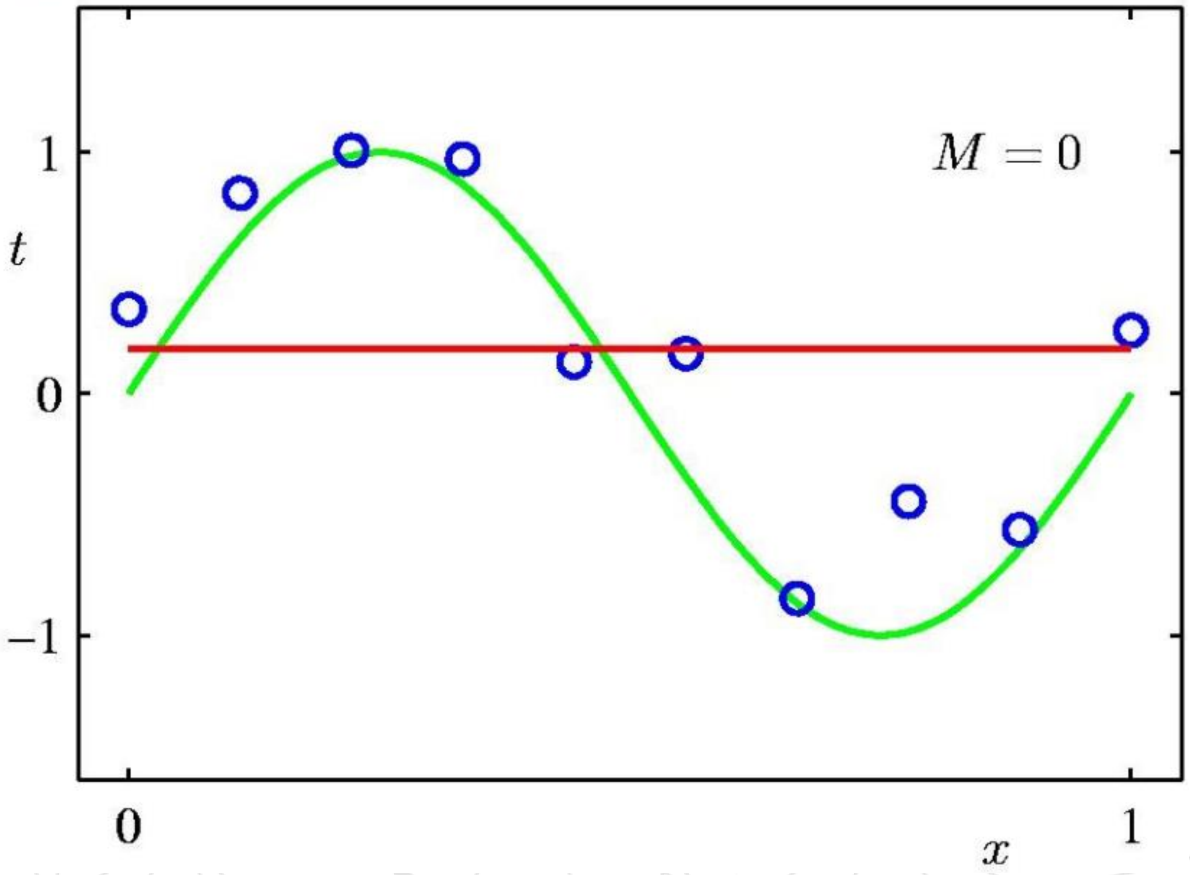
- Данные: sin(x) + случайный шум

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



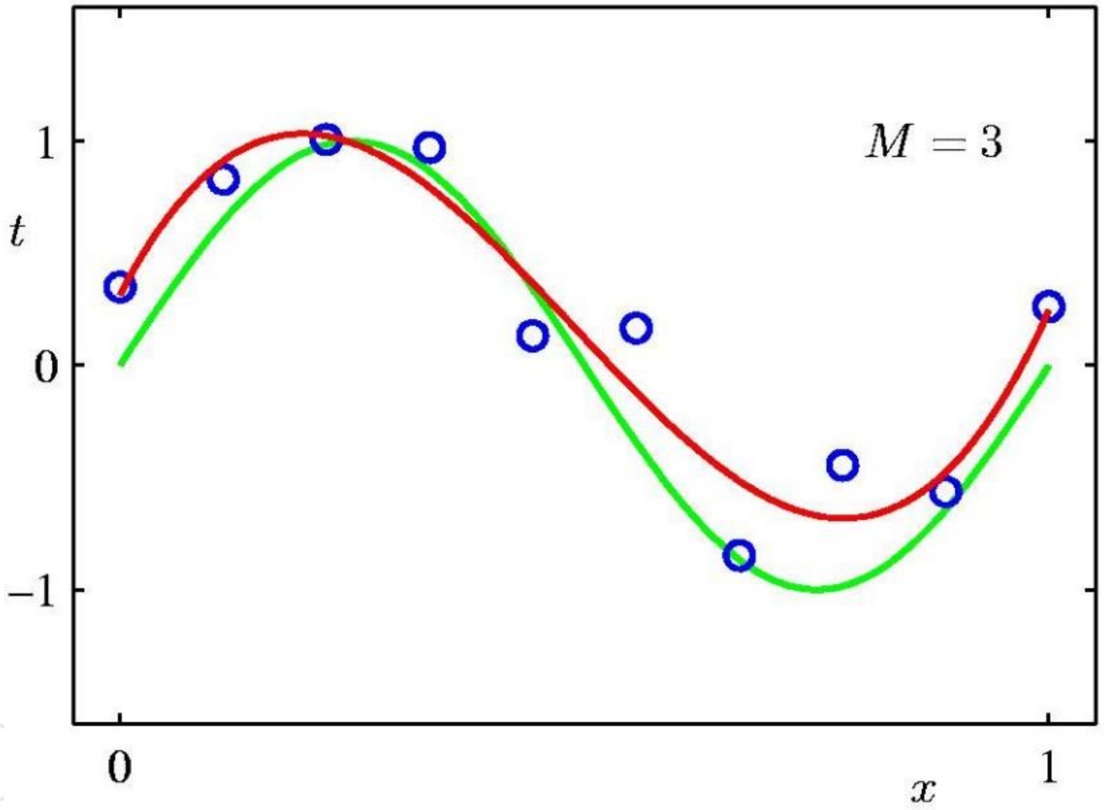
Многочлен нулевой степени M=0

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



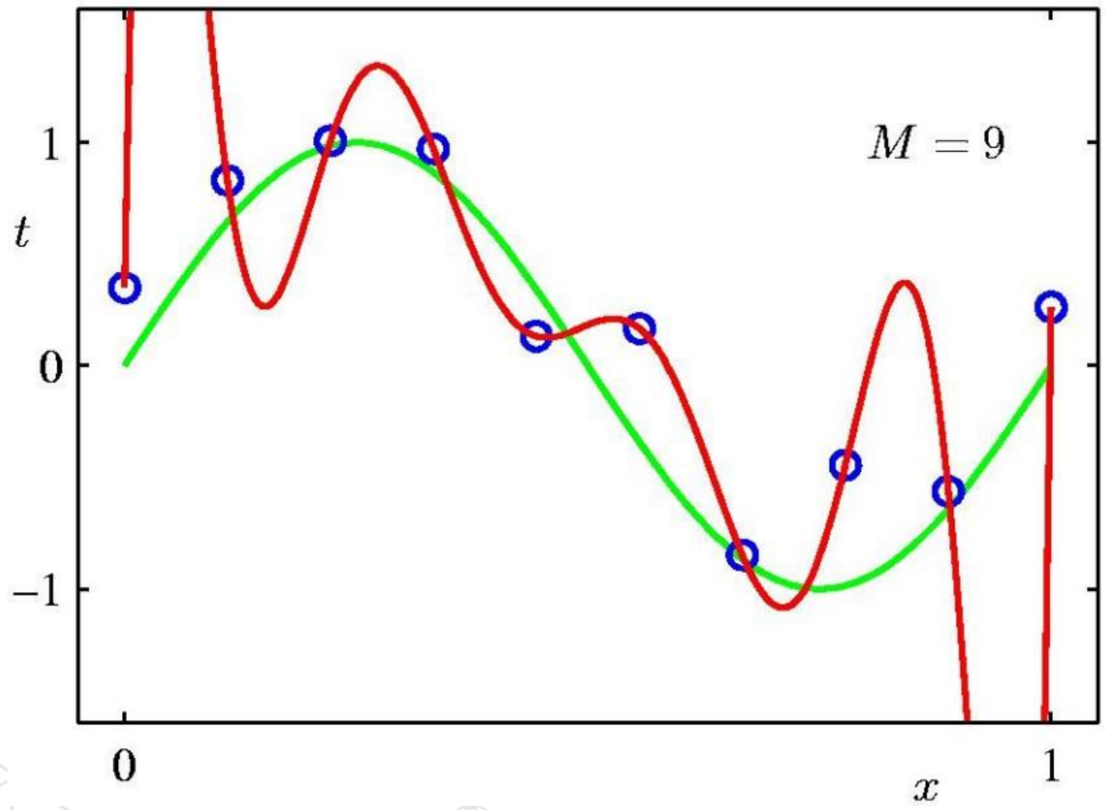
Многочлен степени M=3

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



Многочлен степени M=9

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



Гребневая регрессия: Ridge regression

Штраф за увеличение нормы вектора весов $\|\alpha\|$:

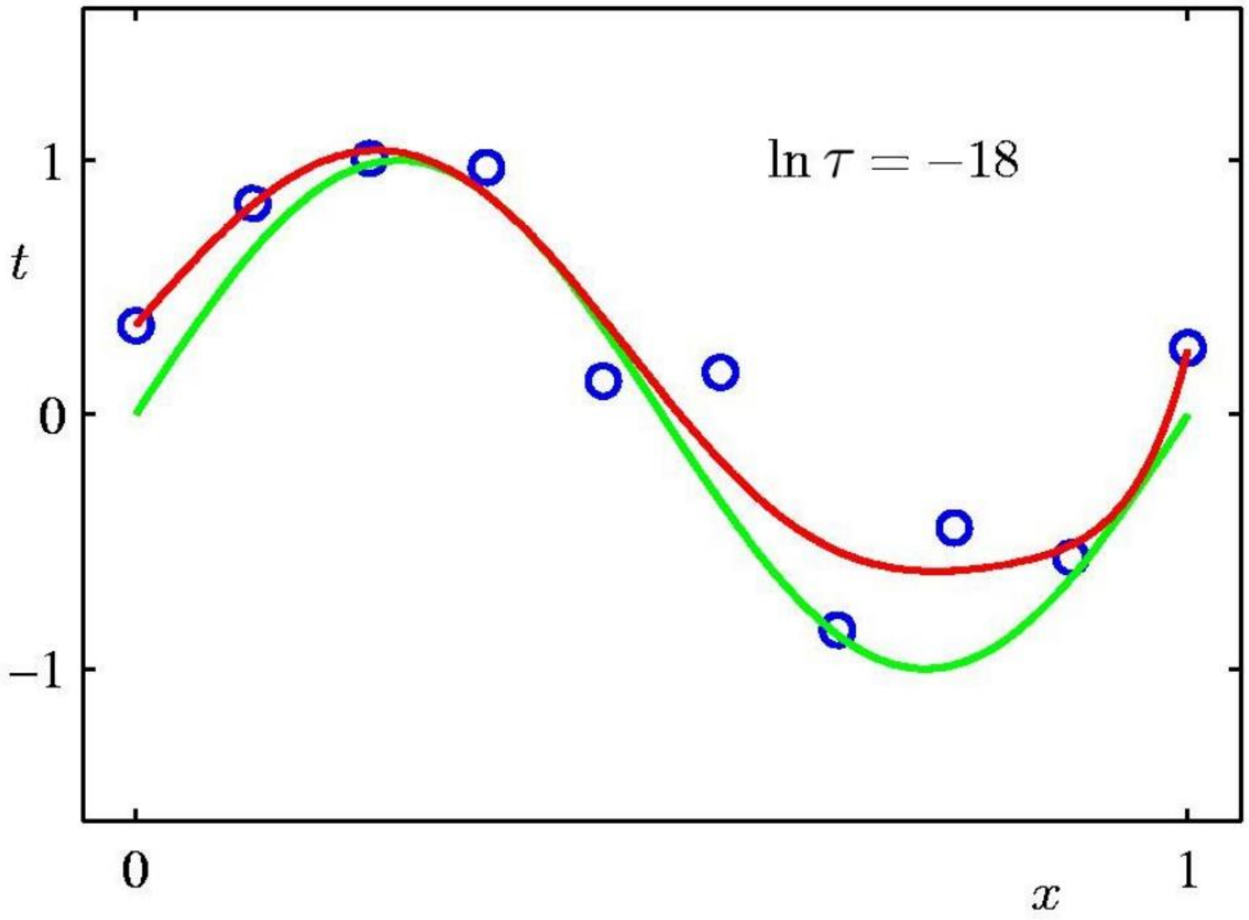
$$Q_\tau(\alpha) = \|F\alpha - y\|^2 + \frac{1}{\sigma}\|\alpha\|^2$$

Модифицированное МНК-решение $\tau = {}^1/_\sigma$
(τI_n — «гребень»):

$$\alpha_\tau^* = (F^\intercal F + \tau I_n)^{-1} F^\intercal y$$

Многочлен 9 степени RidgeRegression

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M = \sum_{j=0}^{M} w_j x^j$$



Сингулярное разложение

Произвольная $\ell \times n$-матрица представима в виде *сингулярного разложения* (singular value decomposition, SVD):

$$F = VDU^{\mathsf{T}}.$$

Основные свойства сингулярного разложения:

1. $\ell \times n$-матрица $V = (v_1, \dots, v_n)$ ортогональна, $V^{\mathsf{T}} V = I_n$, столбцы v_j — собственные векторы матрицы FF^{T};
2. $n \times n$-матрица $U = (u_1, \dots, u_n)$ ортогональна, $U^{\mathsf{T}} U = I_n$, столбцы u_j — собственные векторы матрицы $F^{\mathsf{T}} F$;
3. $n \times n$-матрица D диагональна, $D = \mathrm{diag}\big(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}\big)$, $\lambda_j \geqslant 0$ — собственные значения матриц $F^{\mathsf{T}} F$ и FF^{T}.

Решение МНК через сингулярное разложение

$$\alpha^* = (F^{\mathsf{T}}F)^{-1}F^{\mathsf{T}}y = F^{+}y$$

$$F^{+} = (UDV^{\mathsf{T}}VDU^{\mathsf{T}})^{-1}UDV^{\mathsf{T}} = UD^{-1}V^{\mathsf{T}} = \sum_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}} u_j v_j^{\mathsf{T}}$$

$$\alpha^* = F^{+}y = UD^{-1}V^{\mathsf{T}}y = \sum_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}} u_j (v_j^{\mathsf{T}} y)$$

$$F\alpha^* = P_F y = (VDU^{\mathsf{T}})UD^{-1}V^{\mathsf{T}}y = VV^{\mathsf{T}}y = \sum_{j=1}^{n} v_j (v_j^{\mathsf{T}} y)$$

$$\|\alpha^*\|^2 = \|D^{-1}V^{\mathsf{T}}y\|^2 = \sum_{j=1}^{n} \frac{1}{\lambda_j} (v_j^{\mathsf{T}} y)^2$$

USETECH

Проблема мультиколлинеарности

Число обусловленности n×n-матрицы Σ

$$\mu(\Sigma) = \|\Sigma\|\|\Sigma^{-1}\| = \frac{\max\limits_{u:\ \|u\|=1} \|\Sigma u\|}{\min\limits_{u:\ \|u\|=1} \|\Sigma u\|} = \frac{\lambda_{\max}}{\lambda_{\min}}$$

При умножении обратной матрицы на вектор, $z = \Sigma^{-1}u$, относительная погрешность усиливается в $\mu(\Sigma)$ раз

$$\frac{\|\delta z\|}{\|z\|} \leqslant \mu(\Sigma)\frac{\|\delta u\|}{\|u\|}$$

Проблема мультиколлинеарности

Если матрица $\Sigma = F^T F$ плохо обусловлена, то:

1. решение становится неустойчивым и неинтерпретируемым, $||\alpha^*||$ велико;
2. возникает переобучение:

   на обучении $Q(\alpha^*, X^\ell) = \|F\alpha^* - y\|^2$ мал

   на контроле $Q(\alpha^*, X^k) = \|F'\alpha^* - y'\|^2$ велик

Стратегии устранения мультиколлинеарности и переобучения:

- регуляризация
- отбор признаков
- преобразование признаков

USETECH

Регуляризация с точки зрения SVD-разложения

$$\alpha^*_\tau = (F^\intercal F + \tau I_n)^{-1} F^\intercal y$$

$$\alpha^*_\tau = U(D^2 + \tau I_n)^{-1} D V^\intercal y = \sum_{j=1}^{n} \frac{\sqrt{\lambda_j}}{\lambda_j + \tau} u_j (v_j^\intercal y)$$

Без регуляризации:

$$\alpha^* = F^+ y = U D^{-1} V^\intercal y = \sum_{j=1}^{n} \frac{1}{\sqrt{\lambda_j}} u_j (v_j^\intercal y)$$

Отбор признаков

LASSO — Least Absolute Shrinkage and Selection Operator

$$\begin{cases} Q(\alpha) = \|F\alpha - y\|^2 \to \min\limits_{\alpha}; \\ \sum\limits_{j=1}^{n} |\alpha_j| \leqslant \varkappa; \end{cases}$$

Чем меньше $\varkappa$, тем больше нулевых α

Lasso vs Ridge

LASSO: Чем меньше $\varkappa$, тем больше нулевых α
Ridge: коэффициенты уменьшает, но не делает их нулевыми!

Спасибо за внимание!

Вопросы?