Восстановление регрессии с использованием алгоритма Kernel RLS (KRLS)

Казакевич Анна Юрьевна

24 апреля 2025 г.

Содержание

Постановка задачи: функция предсказания

$$f(x) = \langle \omega, \phi(x) \rangle,$$
$$\omega = (w^T, b), \quad \phi(x) = (\phi^T, 1)^T$$

Цель - минимизация функции потерь вида:

$$L(w) = \sum_{i=1}^{l} (f(x_i) - y_i)^2 = \left\| \Phi_t^\top \omega - y_t \right\|^2, \quad \text{где } y_t = (y_1, \ldots, y_t)^T.$$

$$w_t = \sum_{i=1}^{l} \alpha_i \Phi(x_i) = \Phi_t \alpha_t, \quad \alpha = (\alpha_1, \ldots, \alpha_l)^\top.$$

Процедура построения разреженного решения

- Текущий словарь: $D_{t-1} = (\tilde{x}_j)_{j=1}^{m_{t-1}}$
- Соответствующие векторы в признаковом пространстве: $\Phi(\tilde{x}_j)_{j=1}^{m_{t-1}}$
- Коэффициенты: $a_t \in \mathbb{R}^{m_{t-1}}$

Условие приближенной линейной зависимости (ALD):

$$\delta_t \le \nu, \quad \text{где} \quad \delta_t = \min_{a_t} \left\| \sum_{j=1}^{m_{t-1}} a_{tj} \Phi(\tilde{x}_j) - \Phi(x_t) \right\|^2$$

Дополнительные определения

Матрица коэффициентов:

$$A_t \text{ — матрица коэффициентов}, \quad [A_t]_{ij} = a_{ij}$$

$$[A_t]_{ij} = 0 \quad \text{для } j > m_t.$$

Приближение в пространстве признаков:

$$\phi(x_i) = \sum_{j=1}^{m_t} a_{ij}\phi(\tilde{x}_j) + \phi_i^{\text{res}}, \quad \left(\|\phi_i^{\text{res}}\|^2 \leq \nu\right)$$

$$\phi(x_t) \approx \sum_{j=1}^{m_t} a_{ij}\phi(\tilde{x}_j), \quad \Phi_t \approx \tilde{\Phi}_i A_t^\top$$

Ядерная матрица:

$$\tilde{\Phi}_t = \left[\phi(\tilde{x}_1), \ldots, \phi(\tilde{x}_m)\right], \quad K_t \approx A_t \tilde{K}_t A_t^\top$$

$$K_t = \Phi_t^T \Phi_t$$

Вычисление критерия ALD

Обозначения:

$$[\tilde{K}_{t-1}]_{i,j} = k(\tilde{x}_i, \tilde{x}_j), \quad [\tilde{K}_{t-1}(x_t)]_i = k(\tilde{x}_i, x_t)$$

$$k_{tt} = k(x_t, x_t), \quad i, j = 1, \dots, m_{t-1}.$$

Выражение для δ_t:

$$\delta_t = \min_{a_t} \left\{ a_t^\top \tilde{K}_{t-1} a_t - 2 a_t^\top \tilde{K}_{t-1}(x_t) + k_{tt} \right\}$$

Оптимальные коэффициенты:

$$\tilde{a}_t = \tilde{K}_{t-1}^{-1} \tilde{k}_{t-1}(x_t)$$

Условие ALD:

$$\delta_t = k_{tt} - \tilde{k}_{t-1}(x_t)^\top \tilde{a}_t \leq \nu \tag{1}$$

Новый вид функции потерь

$$w_t = \Phi_t \alpha_t = \Phi_t A_t^\top \alpha_t = \Phi_t \tilde{a}_t$$
$$L(\tilde{\alpha}) = \left\| \Phi_t^\top \Phi_t \tilde{\alpha} - y_t \right\|^2 = \left\| A_t \tilde{K}_t \tilde{\alpha} - y_t \right\|^2$$

Приближения:

$$\Phi_t \sim \tilde{\Phi}_t A_t^T \Rightarrow \Phi_t^T \sim A_t \tilde{\Phi}_t{}^T$$
$$\tilde{\alpha}_t = (A_t \tilde{K}_t)^* y_t = \tilde{K}_t^{-1} A_t^* y_t, \quad A_t^* = (A_t^\top A_t)^{-1} A_t$$

Обновление: Добавление новой точки (ALD не выполнено)

Если $\delta_t > \nu$:

- Добавляем x_t в словарь.
- Обновляем инвертированную матрицу ядровых значений:

$$K_{\tilde{t}}^{-1} = \frac{1}{\delta_t} \begin{pmatrix} \delta_t\, K_{\tilde{t-1}}^{-1} + a_t\, a_t^\top & -a_t \\ -a_t^\top & 1 \end{pmatrix},$$

  где $a_t = K_{\tilde{t-1}}^{-1}\, \mathbf{k}_{t-1}(x_t)$.

- Обновление коэффициентов:

$$\tilde{\alpha}_t = \begin{pmatrix} \tilde{\alpha}_{t-1} - a_t\, \dfrac{e_t}{\delta_t} \\ \dfrac{e_t}{\delta_t} \end{pmatrix}, \quad e_t = y_t - \mathbf{k}_{t-1}(x_t)^\top\, \tilde{\alpha}_{t-1}.$$

Обновление: Без расширения словаря (ALD выполнено)

Если $\delta_t \le \nu$:

- Обновляем вспомогательную матрицу P:

$$P_t = P_{t-1} - \frac{P_{t-1}\, a_t\, a_t^\top\, P_{t-1}}{1 + a_t^\top\, P_{t-1}\, a_t}.$$

- Обновляем коэффициенты:

$$\tilde{\alpha}_t = \tilde{\alpha}_{t-1} + \tilde{K}_{t-1}^{-1}\, q_t\, e_t, \quad q_t = P_{t-1}\, a_t.$$

Псевдокод алгоритма KRLS (ALD)

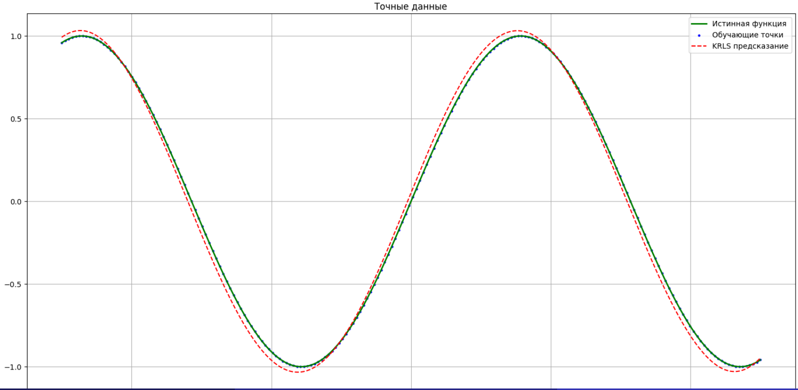
	Cost
Parameter: ν	
Initialize: $\tilde{\mathbf{K}}_1 = [k_{11}]$, $\tilde{\mathbf{K}}_1^{-1} = [1/k_{11}]$,	
$\quad\boldsymbol{\alpha}_1 = (y_1/k_{11})$, $\mathbf{P}_1 = [1]$, $m = 1$	
for $t = 2, 3 \ldots$	
1. Get new sample: $(\mathbf{x}_t, y_t)$	
2. Compute $\tilde{\mathbf{k}}_{t-1}(\mathbf{x}_t)$	$O(m)$
3. ALD test:	
$\boldsymbol{a}_t = \tilde{\mathbf{K}}_{t-1}^{-1}\tilde{\mathbf{k}}_{t-1}(\mathbf{x}_t)$	$O(m^2)$
$\delta_t = k_{tt} - \tilde{\mathbf{k}}_{t-1}(\mathbf{x}_t)^\top \boldsymbol{a}_t$	$O(m)$
if $\delta_t > \nu$ % add $\mathbf{x}_t$ to dictionary	$O(1)$
$\mathcal{D}_t = \{\mathcal{D}_{t-1} \cup \{\tilde{\mathbf{x}}_t\}\}$	$O(1)$
Compute $\tilde{\mathbf{K}}_t^{-1}$ (3.14) with $\tilde{\boldsymbol{a}}_t = \boldsymbol{a}_t$	$O(m^2)$
$\boldsymbol{a}_t = (\mathbf{0}, \ldots, 1)^\top$	$O(m)$
Compute $\mathbf{P}_t$ (3.15)	$O(m)$
Compute $\boldsymbol{\alpha}_t$ (3.16)	$O(m)$
$m := m + 1$	$O(1)$
else % dictionary unchanged	
$\mathcal{D}_t = \mathcal{D}_{t-1}$	
$\mathbf{q}_t = \frac{\mathbf{P}_{t-1}\boldsymbol{a}_t}{1+\boldsymbol{a}_t^\top \mathbf{P}_{t-1}\boldsymbol{a}_t}$	$O(m^2)$
Compute $\mathbf{P}_t$ (3.12)	$O(m^2)$
Compute $\boldsymbol{\alpha}_t$ (3.13)	$O(m^2)$
Output: $\mathcal{D}_t$, $\boldsymbol{\alpha}_t$	

Восстановление регрессии по точным данным

- Генерируется равномерная выборка $x \in [-5, 5]$ с зависимостью:

$$y = \sin(x).$$

- Подобрано $\nu = 0.06125$
- Кол-во опорных векторов = 21
- MSE на тестовой выборке = 0.0017935616860363 32

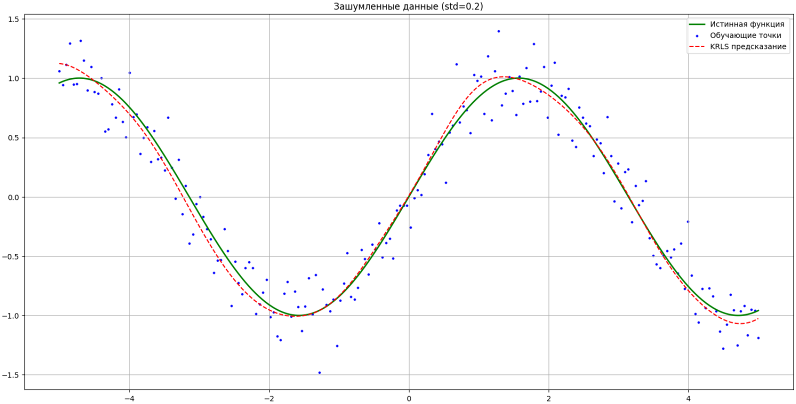


Восстановление регрессии по зашумлённым данным

- Модельные данные с зависимостью:

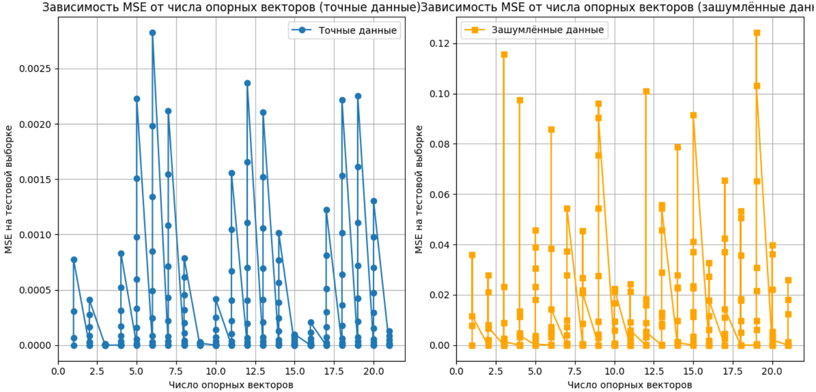
$$y = \sin(x) + \eta, \quad \eta \sim \mathcal{N}(0, 0.2).$$

- Кол-во опорных векторов = 21
- MSE на тестовой выборке = 0.0031202358075215976



Зависимость MSE от числа опорных векторов

- При уменьшении порога ν число опорных точек растёт.
- Снижение MSE происходит до определённого уровня, демонстрируя компромисс между размером модели и точностью.



Полный код алгоритма

Полный код алгоритма, а также более подробный отчёт о методе и его работе можно найти по ссылке
https://github.com/AnsKaz-yu/KRLS-SVM-algorithm/tree/main
Там же в отчёте находится составленный мной подробный псевдокод, который из-за объёма не получилось вставить в презентацию.

Выводы

- Алгоритм KRLS расширяет классический метод RLS для задач регрессии в высокоразмерном пространстве признаков.
- Динамическое формирование разреженного словаря с помощью ALD существенно уменьшает вычислительные затраты.
- Оптимальный выбор порога ν обеспечивает баланс между точностью аппроксимации и эффективностью модели.
- Метод подходит для онлайн-обучения, обновляя модель за фиксированное количество операций.

Литература

1. Engel Y., Mannor S., Meir R. The Kernel Recursive Least Squares Algorithm // Advances in Neural Information Processing Systems (NeurIPS). – 2003.