Московский государственный университет имени М. В. Ломоносова
механико-математический факультет
кафедра вычислительной математики

Старцев В. Ю., 632 группа

Обзор темы (возможной) дипломной работы

Улучшенный метод kNN-регрессии. Улучшенная оценка ошибки аппроксимации градиента многомерной функции по соседним точкам.

г. Москва, 2023 г.

1 Тезисный план.

1.1 Улучшение kNN-регрессии.

Попробуем улучшить kNN-регрессию, заменив аппроксимацию УМО с усреденения значений УМО в k соседних точках:

$$\mathbb{E}(Y|X=x) = \eta(x) \sim \frac{1}{m}\sum_{i=1}^{m}\eta(a_i),$$

на усреденние разложений Тейлора УМО в этих точках:

$$\mathbb{E}(Y|X=x) = \eta(x) \sim \frac{1}{m}\sum_{i=1}^{m}\sum_{k=0}^{n-1}\frac{\nabla^k\eta(a_i)}{k!}(x-a_i)^k.$$

Замечание 1.1. *Задача регрессии - имея тренировочные данные* $(X_1, Y_1), ..., (X_n, Y_n) \in \mathcal{X} \times \mathbb{R}$, *получить такую функцию* $f(x)$, *которая бы по входу* $x \in \mathcal{X}$ *предсказывала метку* $Y \in \mathbb{R}$. *Мерой качества часто служит математическое ожидание квадрата разницы между настоящей и предсказанной метками:*

$$\mathbb{E}[(Y - f(x))^2|X = x].$$

Минимум такого функционала достигается на $f(x) = \eta(x) = \mathbb{E}(Y|X = x)$. *Данным утверждением мотивирована аппроксимация именно УМО.*

Конечная цель: получить оценку на аппроксимацию УМО в точке и показать, что она лучше, чем в случае обыкновенного подхода.

1.2 Аппроксимация градиента многомерной функции в точке по ближайщим соседям.

В рамках задачи необходимо уметь аппроксимировать градиент функции в точках, имея в арсенале значения функции в соседних точках, а также получить оценку на ошибку аппроксимации.

Задача 1.1. *Пусть дана функция* $g : D \subset \mathbb{R}^d \to \mathbb{R}$ *и точка* $a \in D$ *в окрестности которой* $g \in C^{n+1}\mathcal{B}(a)$. *Пусть также в окрестности* a *лежат* m *точек* $v_i = a + h_i\nu_i, i = 1, ..., m$, *где* $h_i = ||v_i - a||$. *Нужно оценить* $\nabla f(a)$.

В силу разложения в ряд Тейлора функции g в окрестности a

$$g(a + h\nu) = g(a) + \sum_{k=1}^{n} h^k\frac{(\nu\cdot\nabla)^k g(a)}{k!} + R_n,$$

где $R_n = \frac{h^{n+1}}{n!}\int_0^1(1-t)^n(\nu\cdot\nabla)^{n+1}g(a+th\nu)dt$, задачу можно решить методом наименьших квадратов. Пусть $\nabla^k = (\frac{\partial^k g}{\partial x_1^k}, ..., \frac{\partial^k g}{\partial x_d^k})^T$. Положим

$$A = \begin{pmatrix} \nu_{11} & \dots & \nu_{1d} & \nu_1^\star \\ \vdots & & \vdots & \vdots \\ \nu_{m1} & \dots & \nu_{md} & \nu_m^\star \end{pmatrix} \in \mathbb{R}^{m\times p},\ q_i = \frac{g(a+h_i\nu_i)-g(a)}{h_i} \in \mathbb{R}^{m\times 1},$$

$$\beta = (\nabla g(a), \nabla g^2(a), ..., \nabla g^n(a))^T \in \mathbb{R}^{p\times 1},$$

где $p = \sum_{s=1}^n C_{s+d-1}^{d-1}$ (количество всех производных порядка s функции d переменных равно количеству способов разложить s шаров по d урнам), а

$$\nu_i^\star = \left(\frac{h_i^{k-1}}{k!}\binom{k}{\alpha_1,...,\alpha_d}\nu_{i1}^{\alpha_1}...\nu_{id}^{\alpha_d}\right)_{2\le k\le n,\ \alpha_1+...+\alpha_d=k}.$$

Тогда имеем $A\beta \approx q$. Решение переопределенной системы находится методом НК (тут замечание об $rk(A) = p, p < m$), в качестве ответа берем первые d координат вектора $\beta:$

$$\nabla\tilde{g}(a) = E_1\tilde{\beta},$$

где $\tilde{\beta} = \arg\min_{\beta\in\mathbb{R}^p}||A\beta - q||_2, E_1 \in \mathbb{R}^{d\times p}-$ первые d строчек единичной матрицы.

Лемма 1.1. *Пусть все производные $n-$ого порядка функции g липщицевы в окрестности точки a: $\frac{\partial^n g}{\partial x_1^{\alpha_1}...\partial x_d^{\alpha_d}} \in Lip_{l_{\alpha_1...\alpha_d}}(\mathcal{B}(a))$ и $l_{max} = \max_{\alpha_1...\alpha_d} l_{\alpha_1...\alpha_d}$. Тогда для любой точки $a+h\nu \in \mathcal{B}(a) : ||\nu||_2 = 1,\ \nu = (\nu_1, ..., \nu_d)^T$ имеем оценку*

$$\left|\sum_{k=1}^n \frac{h^{k-1}}{k!}(\nu\cdot\nabla)^k g(a) - \left[\frac{g(a+h\nu)-g(a)}{h}\right]\right| \le \frac{h^n}{(n+1)!}l_{max}||\nu||_1^n.$$

Доказательство. Для $x = a+h\nu \in \mathcal{B}(a)$ по разложению Тейлора с остаточным членом в интегральной форме имеем:

$$\frac{g(a+h\nu)-g(a)}{h} - \sum_{k=1}^{n-1}\frac{h^{k-1}}{k!}(\nu\cdot\nabla)^k g(a) = \frac{h^{n-1}}{(n-1)!}\int_0^1(1-t)^{n-1}(\nu\cdot\nabla)^n g(a+th\nu)dt.$$

Так как $\int_0^1(1-t)^{n-1}dt = \frac{1}{n}$, прибавим к обеим частям равенства выражение $\frac{h^{n-1}}{(n-1)!}\int_0^1(1-t)^{n-1}(\nu\cdot\nabla)^n g(a)dt = \frac{h^{n-1}}{n!}(\nu\cdot\nabla)^n g(a):$

$$\frac{h^{n-1}}{n!}(\nu\cdot\nabla)^n g(a) - \frac{g(a+h\nu)-g(a)}{h} + \sum_{k=1}^{n-1}\frac{h^{k-1}}{k!}(\nu\cdot\nabla)^k g(a) =$$

$$= \frac{h^{n-1}}{(n-1)!}\int_0^1(1-t)^{n-1}\Big[(\nu\cdot\nabla)^n g(a) - (\nu\cdot\nabla)^n g(a+th\nu)\Big]dt =$$

$$= \frac{h^{n-1}}{(n-1)!} \sum_{\alpha_1+...+\alpha_d=n} \binom{n}{\alpha_1, ..., \alpha_d} \nu_1^{\alpha_1} \cdot ... \cdot \nu_d^{\alpha_d} \int_0^1 (1-t)^{n-1} \left\{ \frac{\partial^n g(a)}{\partial x_1^{\alpha_1} \cdot ... \cdot \partial x_d^{\alpha_d}} - \frac{\partial^n g(a+th\nu)}{\partial x_1^{\alpha_1} \cdot ... \cdot \partial x_d^{\alpha_d}} \right\} dt.$$

Следовательно, получаем оценку

$$\left| \sum_{k=1}^{n-1} \frac{h^{k-1}}{k!} (\nu \cdot \nabla)^k g(a) - \left[\frac{g(a+h\nu) - g(a)}{h} \right] \right| \leq$$

$$\leq \frac{h^{n-1}}{(n-1)!} \sum_{\alpha_1+...+\alpha_d=n} \binom{n}{\alpha_1, ..., \alpha_d} |\nu_1|^{\alpha_1} \cdot ... \cdot |\nu_d|^{\alpha_d} \int_0^1 |1-t|^{n-1} \left| \frac{\partial^n g(a)}{\partial x_1^{\alpha_1} \cdot ... \cdot \partial x_d^{\alpha_d}} - \frac{\partial^n g(a+th\nu)}{\partial x_1^{\alpha_1} \cdot ... \cdot \partial x_d^{\alpha_d}} \right| dt \leq$$

$$\leq \frac{h^{n-1}}{(n-1)!} \sum_{\alpha_1+...+\alpha_d=n} \binom{n}{\alpha_1, ..., \alpha_d} |\nu_1|^{\alpha_1} \cdot ... \cdot |\nu_d|^{\alpha_d} \cdot l_{\alpha_1 ... \alpha_d} \int_0^1 (1-t)^{n-1} ||a - a - th\nu||_2 dt \leq$$

$$\leq \left[\int_0^1 |1-t|^{n-1} || - th\nu||_2 dt = h \int_0^1 |1-t|^{n-1} |t| dt = \frac{h}{n^2+n} \right] \leq \frac{h^n}{(n-1)!} \frac{l_{max}}{n^2+n} (|\nu_1| + ... + |\nu_d|)^n =$$

$$= \frac{h^n}{(n+1)!} l_{max} ||\nu||_1^n$$

□

Теорема 1.1. *В условиях предыдущей леммы для оценки на градиент, полученной методом НК, имеем*

$$||\nabla f(a) - E_1 \tilde{\beta}||_2 \leq \frac{l_{max} h_{max}^n}{\sigma_1 (n+1)!} \sqrt{\sum_{i=1}^m ||\nu_i||_1^{2n}},$$

где $a_i = a + \nu_i h_i, h_{max} = \max_{1 \leq k \leq m} h_k, \sigma_1 -$ *наименьшее сингулярное значение матрицы* A.

Доказательство. Пусть $E_2 \in \mathbb{R}^{(p-d) \times p} -$ матрица из последних $p - d$ строк $p \times p$ единичной матрицы. Тогда

$$||\beta - \tilde{\beta}||_2^2 = ||E_1(\beta - \tilde{\beta})||_2^2 + ||E_2(\beta - \tilde{\beta})||_2^2 \geq ||E_1(\beta - \tilde{\beta})||_2^2 = ||\nabla g(a) - E_1 \tilde{\beta}||_2^2$$

С другой стороны, если $A = U \Sigma V^T -$ SVD-разложение матрицы A имеем

$$||\beta - \tilde{\beta}||_2^2 = ||\beta - V \Sigma^{-1} U^T q||_2^2 = ||V \Sigma^{-1} U^T (U \Sigma V^T \beta - q)||_2^2 \leq ||\Sigma^{-1}||_2^2 ||A\beta - q||_2^2 = \frac{1}{\sigma_1^2} ||A\beta - q||_2^2.$$

Наконец, используя результат **Леммы 1.1**, получим искомую оценку

$$||A\beta - q||_2^2 = \sum_{i=1}^{m} \left| \sum_{k=1}^{n} \frac{h_i^{k-1}}{k!} (\nu_i \cdot \nabla)^k g(a) - \left\{ \frac{g(a + h_i\nu_i) - g(a))}{h_i} \right\} \right| \leq$$

$$\leq \sum_{i=1}^{m} \left(\frac{h_i^n}{(n+1)!} l_{max} ||\nu_i||_1^n \right) \leq \left(\frac{l_{max}}{(n+1)!}\right)^2 (h_{max}^n)^2 \sum_{i=1}^{m} ||\nu_i||_1^{2n}.$$

Следовательно

$$||\nabla g(a) - E_1\tilde{\beta}||_2 \leq ||\beta - \tilde{\beta}||_2 \leq \frac{l_{max} h_{max}^n}{\sigma_1 (n+1)!} \sqrt{\sum_{i=1}^{m} ||\nu_i||_1^{2n}}$$

□

Оказывается, данную оценку можно улучшить. Это связано с двумя следующими утверждениями.

Утверждение 1.1. *Рассмотрим представление матрицы A:* $A = (A_1|A_2)$, *где*

$$A_1 = \begin{pmatrix} \nu_{11} & \dots & \nu_{1d} \\ \vdots & & \vdots \\ \nu_{m1} & \dots & \nu_{md} \end{pmatrix}, \ A_1 \in \mathbb{R}^{m \times d},$$

$$A_2 = \begin{pmatrix} \nu_1^\star \\ \vdots \\ \nu_m^\star \end{pmatrix}, \ A_2 \in \mathbb{R}^{m \times (p-d)}.$$

Пусть $A_2 = Q \begin{pmatrix} A_{12} \\ 0 \end{pmatrix}, A_{12} \in \mathbb{R}^{(p-d)\times(p-d)}$ *– QR-разложение матрицы* A_2 *с верхнетреугольной* A_{12}. *Пусть далее* $Q^T = \begin{pmatrix} Q_1^T \\ Q_2^T \end{pmatrix}$ *и* $Q^T A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix}$, *где* $A_{11} = Q_1^T A_1, \ A_{12} = Q_1^T A_2$ *и* $A_{21} = Q_2^T A_1$. *Тогда решения* $E_1\tilde{\beta}$ *и* $\tilde{\tau} = argmin_{\tau \in \mathbb{R}^d} ||A_{21}\tau - Q_2^T q||_2$ *совпадают и минимумы* $min_{\tau \in \mathbb{R}^d} ||A_{21}\tau - Q_2^T q||_2 = min_{\beta \in \mathbb{R}^p} ||A\beta - q||_2$ *также совпадают.*

Доказательство. Положим $D := A\beta - q$, где

$$\beta = (\beta_1, \beta_2)^T \in \mathbb{R}^{p \times 1}, \beta_1 \in \mathbb{R}^d, \beta_2 \in \mathbb{R}^{p-d}.$$

Имеем,

$$(Q^T A_1 | Q^T A_2) - Q^T q = Q^T D \Rightarrow \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} - \begin{pmatrix} q_1 \\ q_2 \end{pmatrix} = Q^T D.$$

$Q-$ ортогональная, следовательно

$$||D||_2^2 = ||Q^T D||_2^2 = ||A_{11}\beta_1 + A_{12}\beta_2 - q_1||_2^2 + ||A_{21}\beta_1 - q_2||_2^2.$$

Таким образом, в силу единственности решения (полноранговость A) исходной задачи НК, первые p координат вектора $\tilde{\beta} = \arg\min_{\beta \in \mathbb{R}^p} ||A\beta - q||_2$ совпадают с $\tilde{\beta_1} = \arg\min_{\beta_1 \in \mathbb{R}^d} ||A_{21}\beta_1 - q_2||_2$. Так как $A_{12}-$ верхнетреугольная, то система

$$A_{11}\tilde{\beta_1} + A_{12}\beta_2 - q_1 = 0$$

имеет единственное решение $\tilde{\beta_2}$. Значит

$$\tilde{\beta} = (\tilde{\beta_1}, \tilde{\beta_1})^T$$

и минимум норм невязок для обозначенных задач совпадают.

$\square$

Утверждение 1.2. *Пусть σ_1 и $\hat{\sigma}_1$ наименьшие сингулярные значения матриц A и A_{21} соответственно. Тогда $\sigma_1 \leq \hat{\sigma}_1$.*

Доказательство. Так как $\inf_{x \neq 0} \frac{||Ax||_2^2}{||x||_2^2} = \inf_{x \neq 0} \frac{x^T A^T A x}{x^T x} = \sigma_1^2$, то

$$\sigma_1^2 = \inf_{x \neq 0} \frac{||Ax||_2^2}{||x||_2^2} = \inf_{x \neq 0} \frac{||Q^T A x||_2^2}{||x||_2^2} = \inf_{x \neq 0} \frac{|| \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} ||_2^2}{||x_1||_2^2 + ||x_2||_2^2}.$$

Пусть u_1, v_1- соотвественно правый и левый сингулярные векторы, отвечающие наименьшему сингулярному значению $\hat{\sigma}_1$ матрицы A_{21}. Положим $x_1 = v_1, x_2 = -A_{12}^{-1} A_{11} v_1$. Так как $||v_1||_2^2 = 1$, имеем

$$\sigma_1^2 \leq \frac{\hat{\sigma}_1^2}{1 + ||x_2||^2},$$

откуда окончательно

$$\sigma_1 \leq \frac{\hat{\sigma}_1}{\sqrt{1 + ||x_2||^2}} \leq \hat{\sigma}_1.$$

$\square$

В качестве следствия получаем улучшенную оценку на ошибку аппроксимации градиента:

Следствие 1.1.

$$||\nabla g(a) - E_1 \tilde{\beta}||_2 \leq \frac{l_{max} h_{max}^n}{\hat{\sigma}_1 (n+1)!} \sqrt{\sum_{i=1}^m ||\nu_i||_1^{2n}} \leq \frac{l_{max} h_{max}^n}{\sigma_1 (n+1)!} \sqrt{\sum_{i=1}^m ||\nu_i||_1^{2n}}$$

Доказательство.

$$||\nabla g(a) - E_1\tilde{\beta}||_2 = ||\nabla g(a) - A_{21}^{\dagger}Q_2^T q||_2 \text{ (вследствие Утверждения 1.1)}$$

$$= ||A_{21}^{\dagger}(A_{21}\nabla g(a) - Q_2^T q)||_2 \leq \frac{1}{\hat{\sigma}_1}||Q_2^T(A_1\nabla g(a) - q)||_2 = \frac{1}{\hat{\sigma}_1}||\, Q_2^T[(A_1|A_2)\beta - q]\,||_2$$

$$\leq \frac{1}{\hat{\sigma}_1}||A\beta - q||_2 \leq \frac{l_{max}h_{max}^n}{\hat{\sigma}_1(n+1)!}\sqrt{\sum_{i=1}^{m}||\nu_i||_1^{2n}} \text{ (см. док-во Теоремы 1.1).}$$

В силу Утверждения 1.2 получаем требуемый результат. □

Данная теория легко дополняется, если вклад соседей взвесить (например, соответственно их расстояниям до тестовой точки).

Замечание 1.2. *Пусть* $W = diag(w_1, ..., w_m) \in \mathbb{R}^{m\times m}$, $w_i \geq 0$, $w_{max} = \max_{i=1,...,m} w_i$ *- диагональная матрица весов;* $\overline{A} = WA$, *взвешенная матрица для задачи НК;* $\overline{\sigma}_1-$ *минимальное ненулевое сингулярное значение матрицы* $\overline{A}$. *Пусть для функции* g *выполнены условия* ***Леммы 1.1***. *Тогда для* $\tilde{\beta} = \arg\min_{\beta\in\mathbb{R}^p}||\overline{A}\beta - Wq||_2$ *имеет место оценка*

$$||\nabla g(a) - E_1\tilde{\beta}||_2 \leq \frac{l_{max}h_{max}^n w_{max}}{\overline{\sigma}_1(n+1)!}\sqrt{\sum_{i=1}^{m}||\nu_i||_1^{2n}}.$$

Так как $||W||_2 = w_{max}$, $\kappa(W) = ||W|| \cdot ||W^{-1}||$ *и* $\sigma_1 = \inf_{||x||=1}||Ax|| \leq ||W^{-1}||\inf_{||x||=1}||\overline{A}x|| = ||W^{-1}||\overline{\sigma}_1$, *то оценку можно переписать в терминах числа обусловленности матрицы* W :

$$||\nabla g(a) - E_1\tilde{\beta}||_2 \leq \frac{l_{max}h_{max}^n\kappa(W)}{\sigma_1(n+1)!}\sqrt{\sum_{i=1}^{m}||\nu_i||_1^{2n}}.$$

Доказательство. Пусть $\overline{U\Sigma V}^T$ - SVD-разложение матрицы $\overline{A}$. Аналогично **Теореме 1.1** имеем

$$||\beta - \tilde{\beta}||_2^2 \geq ||\nabla g(a) - E_1\overline{V\Sigma}^{-1}\overline{U}^T Wq||_2^2$$

и

$$||\beta - \tilde{\beta}||_2^2 = ||\beta - \overline{V\Sigma}^{-1}\overline{U}^T Wq|| \leq ||\overline{\Sigma}^{-1}||_2^2||\overline{A}\beta - Wq||_2^2 \leq \frac{1}{\overline{\sigma}_1^2}||W||_2^2||A\beta - q||_2^2 =$$

$$= \frac{1}{\overline{\sigma}_1^2}w_{max}^2||A\beta - q||_2^2$$

Оценку для $||A\beta - q||_2^2$ берем из доказательства **Теоремы 1.1** □

2 Список литературы

1. Гамильтон У. Р. Избранные труды: оптика, динамика, кватернионы. Серия «Классики науки». М., Наука, 1994.
2. Клейн Ф. Лекции о развитии математики в XIX столетии. М., Наука. 1989.
3. Maxwell J.C. A treatise on electricity and magnetism. London, 1873. Vol. 2.