

# Анализ свойств ансамбля локально аппроксимирующих моделей

Исламов Рустем Ильфакович

Московский физико-технический институт

Консультант Грабовой А. В.

Эксперт Стрижов В. В.

30 апреля 2020 г

# Цель работы

## Задача

Аппроксимация данных, в которых данные порождены несколькими источниками

## Проблема

Выборки, порожденные несколькими источниками затруднительно аппроксимировать одной моделью

## Предлагаемое решение

Построение универсальной модели в виде ансамбля локальных моделей

# Постановка задачи построения ансамбля

## Данные

Матрица признаков  $\mathbf{X} \in \mathbb{R}^{N \times n}$  и выборка данных  $\mathfrak{D} = \{(\mathbf{x}_i, y_i) : i \in \mathcal{I}\}$ .

## Гипотеза порождения данных

Данные порождены  $K$  источниками. Это предположение индуцирует разбиение множества индексов  $\mathcal{I}$  на  $K$  непересекающихся подмножеств  $\mathcal{I}_k$ :

$$\mathcal{I} = \bigsqcup_{k=1}^K \mathcal{I}_k.$$

Разбиение  $\mathcal{I}$  индуцирует разбиение выборки данных  $\mathfrak{D}$  и множества объектов  $\Omega$ :

$$\mathfrak{D} = \bigsqcup_{k=1}^K \mathfrak{D}_k, \quad \Omega = \bigsqcup_{k=1}^K \Omega_k.$$

## Определение

Модель  $g_k$  называется локальной, если она аппроксимирует подвыборку

$$\mathfrak{D}_k = \{(\mathbf{x}_i, y_i) : i \in \mathcal{I}_k\}.$$

В данной работе каждая локальная модель является линейной. Локальные модели объединены в ансамбль локальных моделей.

## Определение

Ансамбль локальных моделей — мультимодель, определяющая правдоподобие веса  $\pi_k$  каждой локальной модели  $g_k$  на признаковом описании объекта  $\mathbf{x}$ .

$$f = \sum_{k=1}^K \pi_k g_k(\mathbf{x}, \mathbf{w}_k), \quad \pi_k(\mathbf{x}, \mathbf{V}) : \mathbb{R}^{n \times |\mathbf{V}|} \rightarrow [0, 1], \quad \sum_{k=1}^K \pi_k(\mathbf{x}, \mathbf{V}) = 1,$$

где  $f$  — ансамбль локальных моделей,  $g_k$  — локальная модель,  $\pi_k$  — шлюзовая функция,  $\mathbf{V}$  — параметры шлюзовой функции.

### Определение

Расстояние между локальными моделями равно выборочному коэффициенту корреляции Пирсона на выборке  $\mathbf{X}$  и вычисляется по формуле

$$\rho(g_i, g_j) = \frac{\sum_{l=1}^N (X_{il} - \bar{X}_i) (X_{jl} - \bar{X}_j)}{\sqrt{\sum_{l=1}^N (X_{il} - \bar{X}_i)^2 \sum_{l=1}^N (X_{jl} - \bar{X}_j)^2}},$$

$X_{il} = (\mathbf{X}\mathbf{w}_i)_l$ ,  $X_{jl} = (\mathbf{X}\mathbf{w}_j)_l$ ,  $\bar{X}_i = \frac{1}{N} \sum_{l=1}^N X_{il}$ ,  $\bar{X}_j = \frac{1}{N} \sum_{l=1}^N X_{jl}$ ,  $\mathbf{w}_i, \mathbf{w}_j$  — параметры локальных моделей  $g_i, g_j$ .

Данные первого эксперимента

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}.$$

$$y_m = \alpha_m x_m + \varepsilon, \quad \varepsilon \in \mathcal{N}(0, 1).$$

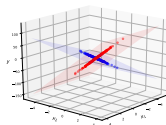


Рис.: Линейная модель.

Данные второго эксперимента

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1 & \varepsilon_1 \\ \varepsilon_2 & x_2 \end{pmatrix}.$$

$$y_m = \alpha_m x_m + \varepsilon, \\ \varepsilon_{1,2} \in \mathcal{N}(0, 1) \quad \varepsilon \in \mathcal{N}(0, 1).$$

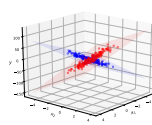
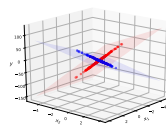
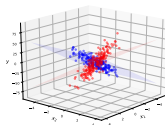


Рис.: Ансамбль локальных моделей.

# Исследование расстояния между моделями

Данные эксперимента  $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ ,  $\hat{\mathbf{X}} = \begin{pmatrix} x_1 & \varepsilon_1 \\ \varepsilon_2 & x_2 \end{pmatrix}$ , где  $\varepsilon_1, \varepsilon_2 \in \mathcal{N}(0, \sigma)$ .

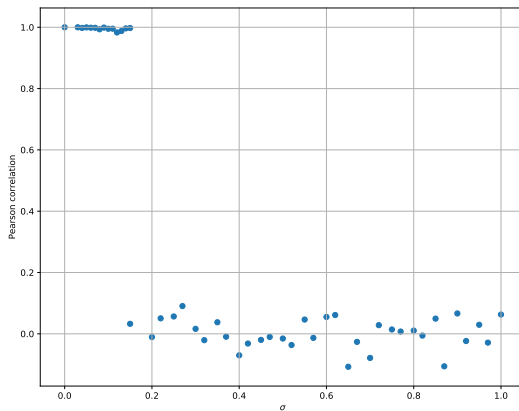


Рис.: График зависимости расстояния между локальными моделями от шума.

## Полученные результаты

- Качество аппроксимации выборки, порожденной несколькими источниками, при использовании ансамбля локальных моделей выше, чем у одной модели
- При увеличении параметра шума модели становятся независимыми



- Исследование оптимального количества локальных моделей в ансамбле
- Использование расстояния между локальными моделями как регуляризатор

- ① *V. V. Strijov A. V. Grabovoy.* Prior distribution choices for a mixture of experts. Machine learning and data analysis, 2020.
- ② *Christopher M. Bishop.* Pattern Recognition and Machine Learning. SPRINGER NATURE, 2011.
- ③ *Павлов К.В.* Выбор многоуровневых моделей в задачах классификации, 2012
- ④ *Esen Y.S., Wilson J., Gader P.D.* Twenty Years of Mixture of Experts. IEEE Transactions on Neural Networks and Learning Systems. 2012. Issues. 23. No 8. P. 1177-1193.