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 — анализ свойств ансамбля локальных моделей

Задача

Аппроксимация данных, в которых данные порождены несколькими источниками

Предлагаемое решение

Построение универсальной модели в виде ансамбля локальных моделей

Требования к универсальному аппроксиматору

- Локальный модели должны быть простыми
- Локальные модели должны аппроксимировать разные подмножества объектов

Список литературы

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.

Постановка задачи построения ансмабля

Выборка

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

Гипотеза порождения данных

Выборка порождена K источниками. Это предположение индуцирует разибение 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 Ω:

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

Определение

Модель g_k называется локальной, если она аппроксимирует подвыборку

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

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

Определение

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

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

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

Расстояние между локальными моделями

Определение

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

$$\rho(\mathrm{g}_i, \mathrm{g}_j) = \frac{\sum\limits_{l=1}^{N} \left(\mathrm{X}_{il} - \overline{\mathrm{X}}_i\right)\left(\mathrm{X}_{jl} - \overline{\mathrm{X}}_j\right)}{\sqrt{\sum\limits_{l=1}^{N} \left(\mathrm{X}_{jl} - \overline{\mathrm{X}}_j\right)^2 \sum\limits_{l=1}^{N} \left(\mathrm{X}_{jl} - \overline{\mathrm{X}}_j\right)^2}},$$

$\mathrm{X}_{il} = (\mathbf{X}\mathbf{w}_i)_l, \mathrm{X}_{jl} = (\mathbf{X}\mathbf{w}_j)_l, \overline{\mathrm{X}}_i = \frac{1}{N}\sum\limits_{l=1}^{N} \mathrm{X}_{il}, \overline{\mathrm{X}}_i = \frac{1}{N}\sum\limits_{l=1}^{N} \mathrm{X}_{il}, \mathbf{w}_i, \mathbf{w}_j$ — параметры локальных моделей g_i, g_j.

Цель введения расстояния

Локальные модели должны быть независимыми, расстояние между ними должно быть близко к нулю.

Базовый эксперимент

Цель базового эксперимента

Показать, что выборка, порожденная несколькими источниками, плохо аппроксимируется одной моделью.

Данные первого эксперимента

$$\mathbf{y} = \begin{pmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathrm{x}_1 & 0 \\ 0 & \mathrm{x}_2 \end{pmatrix}.$$

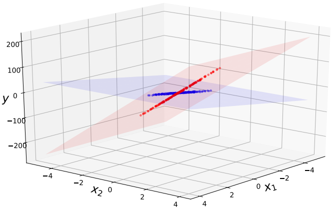
$$\mathrm{y}_m = \alpha_m \mathrm{x}_m + \varepsilon, \qquad \varepsilon \in \mathcal{N}(0, 1).$$

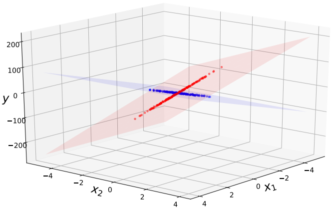
Данные второго эксперимента

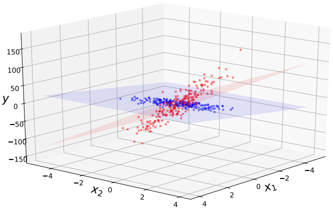
$$\mathbf{y} = \begin{pmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} \mathrm{x}_1 & \varepsilon_1 \\ \varepsilon_2 & \mathrm{x}_2 \end{pmatrix}.$$

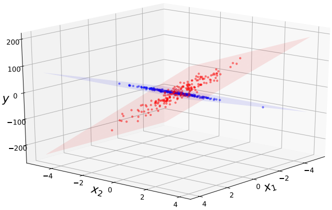
$$\mathrm{y}_m = \alpha_m \mathrm{x}_m + \varepsilon,$$

$$\varepsilon_{1,2} \in \mathcal{N}(0, 1) \qquad \varepsilon \in \mathcal{N}(0, 1).$$





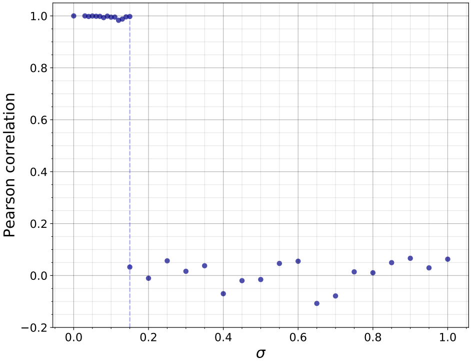




Ансамбль локальных моделей лучше аппроксимирует выборку, порожденную несколькими источниками.

Исследование расстояния между моделями на синтетической выборке

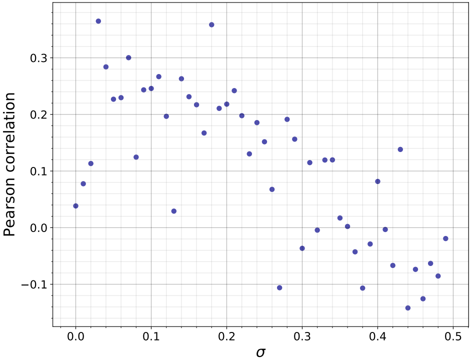
Выборка для эксперимента $\mathbf{y} = \begin{pmatrix} \mathrm{y}_1 \\ \mathrm{y}_2 \end{pmatrix}, \hat{\mathbf{X}} = \begin{pmatrix} \mathrm{x}_1 & \varepsilon_1 \\ \varepsilon_2 & \mathrm{x}_2 \end{pmatrix}$, где $\varepsilon_1, \varepsilon_2 \in \mathcal{N}(0, \sigma)$.



При параметре шума σ меньшем порогового значения локальные модели практически одинаковы, при σ большем порога локальные модели независимы.

Исследование расстояния между моделями на реальных выборках

Выборка для эксперимента $\tilde{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_b \\ \mathbf{y}_s \end{pmatrix} \in \mathbb{R}^{673}, \qquad \tilde{\mathbf{X}} = \begin{pmatrix} & \mathbf{X}_b \\ \mathbf{X}_s & \mathcal{E} \end{pmatrix} \in \mathbb{R}^{673 \times 13}.$



При увеличении параметра шума есть тенденция к уменьшению расстояния между локальными моделями, они становятся более независимыми.

Заключение

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

Дальнейшие исследования

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