Разработка эффективных методов построения ансамблевых регрессионных линейных моделей, основанных на минимизации ошибки

Александр Сергеевич Суржанский

Московский Государственный Университет

13 декабря 2024 г.

Постановка задачи

- Предположим, что у нас имеется набор из l предикторов, прогнозирующих значения некоторой переменной Y
- Прогноз, вычисляемый i-м предиктором для некоторого объекта ω, далее будет обозначаться через $z_i(\omega)$
- Пусть $c = (c_1, \dots, c_l)$ – вектор действительных неотрицательных коэффициентов, удовлетворяющий условию $\sum_{i=1}^{l} c_i = 1$
- Будем рассматривать выпуклые корректирующие процедуры (ВКП), вычисляющие коллективный прогноз $Z(\omega, c)$ в виде

$$Z(\omega, c) = \sum_{i=1}^{l} c_i z_i(\omega)$$

Квадратичная ошибка

При прогнозировании значения Y для объекта ω квадратичная ошибка ВКП может быть описана в форме разности:

$$[Y(\omega)-Z(\omega,c)]^2=\sum_{i=1}^{l}c_i[Y(\omega)-z_i(\omega)]^2-\sum_{i=1}^{l}c_i[z_i(\omega)-Z(\omega,c)]^2$$

Задачу выбора оптимальной ВКП можно рассматривать как задачу минимизации ожидаемой ошибки на пространстве Ω всех потенциально прогнозируемых объектов генеральной совокупности:

$$\Delta=\mathbb{E}_\Omega\left(\sum_{i=1}^{l}c_i[Y(\omega)-z_i(\omega)]^2-\sum_{i=1}^{l}c_i[z_i(\omega)-Z(\omega,c)]^2\right)$$

Задача квадратичного программирования

Обозначим:

- $\delta_i = \mathbb{E}_\Omega[Y(\omega) - z_i(\omega)]^2$ - математическое ожидание ошибки индивидуального прогностического алгоритма
- $\rho_{ij} = \mathbb{E}_\Omega[z_i(\omega) - z_j(\omega)]^2$ - величина, характеризующая степень расхождения i-го и j-го прогностических алгоритмов

Задачу минимизации обобщенной ошибки можно свести к задаче квадратичного программирования:

$$\begin{cases} \sum_{i=1}^{l} c_i\delta_i - \frac{1}{2}\sum_{i,j=1}^{l} c_i c_j \rho_{ij} \to \min \\ \sum_{i=1}^{l} c_i = 1; c_i \geq 0, i = 1, \dots l \end{cases}$$

Нескоратимые предикторы

Под условием несократимости набора предикторов понимается условие невозможности удаления из набора какого-либо элемента без уменьшения точности соответствующей оптимальной ВКП.
Пусть:

$$\overline{D}_l = \{c | \sum_{i=1}^{l} c_i = 1; c_i \geq 0, i = 1, \dots, l\}$$

$$D_l = \{c | \sum_{i=1}^{l} c_i = 1; c_i > 0, i = 1, \dots, l\}$$

Набор предикторов будет называться несократимым, если существует точка $c \in D_l$ такая, что

$$\Delta(c) < \Delta(c') \qquad \forall c' \in \overline{D}_l \backslash D_l$$

Несократимость в случае двух предикторов

Для того чтобы обобщенная квадратичная ошибка оптимальной ВКП по двум предикторам была меньше, чем обобщенная квадратичная ошибка каждого из предикторов, необходимо и достаточно, чтобы выполнялось неравенство

$$|\delta_2 - \delta_1| < \rho_{12}$$

В этом случае обобщенная квадратичная ошибка достигает своего минимума, если

$$c_1 = \frac{\rho_{12} + \delta_2 - \delta_1}{2\rho_{12}}$$

$$c_2 = \frac{\rho_{12} + \delta_1 - \delta_2}{2\rho_{12}}$$

Несократимость в общем случае

Одновременное выполнение следующих условий является необходимым и достаточным условием несократимости набора предикторов:

$$\sum_{i=1}^{l}[\delta_i\overline{\rho}_{ij} + \frac{\frac{1}{2} - \sum_{k,t=1}^{l}\delta_t\overline{\rho}_{kj}}{\sum_{k,t}^{l}\overline{\rho}_{kt}}\overline{\rho}_{ij}] > 0,$$

где матрица $||\overline{\rho}_{ij}||_{l\times l}$ является обратной матрице $||\rho_{ij}||_{l\times l}$, и положительной определенности соответствующей квадратичной формы

$$-\frac{1}{2}\sum_{i,j=1}^{l}\rho_{ij}z_iz_j$$

для любого вещественного вектора $z_1, \ldots, z_l$ такого, что $\sum_{i=1}^{l} z_i = 0$

Свойства ВКП

Важным свойством ВКП является снижение дисперсии коллективных прогнозов $Z(\omega)$ по отношению к средней дисперсии по ансамблю, так как выполнено следующее утверждение:

$$\sqrt{\mathbb{D}(Z)} \leq \sum_{i=1}^{l} c_i \sqrt{\mathbb{D}(z_i)}$$

Описание алгоритма

1. Базовые предикторы: $z_i(\omega) = \alpha_i + \beta_i X_i(\omega) \quad i = 1, \dots, l$
2. Оценка параметров δ_i и ρ_{ij} с помощью метода leave-one-out
3. Выбор пары предикторов с максимальным расстоянием ρ_{ij}
4. Поиск оптимальной ВКП:

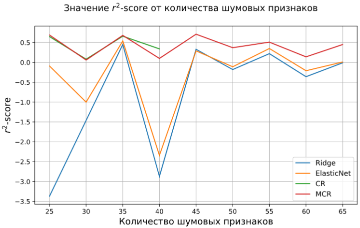
$$Z(\omega, c^0) = \sum_{i=1}^{l} c_i^0 \alpha_i + \sum_{i=1}^{l} c_i^0 \beta_i X_i(\omega)$$

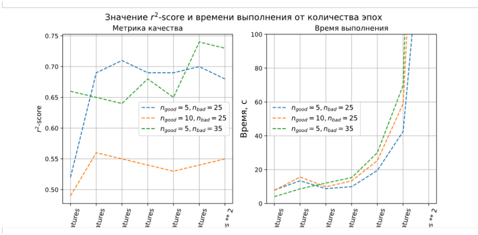
5. Усреднение полученных коэффициентов для каждой выпуклой комбинации предикторов
6. Повторение шагов (1) - (5) в течении заранее заданного числа эпох
7. Добавление линейного преобразования для $Z(\omega)$:

$$Y_{\text{pred}} = \alpha + \beta Z$$

Эксперименты на синтетических данных

- При увеличении числа шумовых признаков модифицированный метод не имеет существенных потерь в качестве, для всех экспериментов значения r^2-score у MCR превышали значения у Ridge и ElasticNet





- Увеличение числа эпох не влечёт за собой обязательное улучшение качества предсказаний метода. Оптимальным значением параметра по r^2-score и времени выполнения является общее количество признаков в выборке.

Эксперименты на реальных данных

Предлагается спрогнозировать определённый параметр химического элемента на основании его набора характеристик

Модель	**Датасет 1**		**Датасет 2**	
	r^2**-score**	**Ненулевые признаки**	r^2**-score**	**Ненулевые признаки**
ВКП	0.9367	19	0.9521	31
СМС	0.9527	42	0.9613	68
Ridge	0.8903	83	0.9511	87
Lasso	0.8427	39	0.9492	71
ElasticNet	0.8849	62	0.9517	80

- Главным преимуществом предложенного алгоритма является способность существенного отбора признаков без значительных потерь в качестве

Основные пункты

- Поиск коэффициентов оптимального ВКП сводится к задаче квадратичного программирования, которая решается в терминах избыточности набора предикторов
- Установлены необходимые и достаточные условия избыточности для наборов из двух предикторов, а также из произвольного числа предикторов
- Продемонстрирована способность естественного отбора признаков без значительных потерь в качестве

Дальнейшие исследования

- В будущем алгоритм можно усовершенствовать, используя концепцию дивергентного леса