

Предсказание временных рядов с помощью градиентного бустинга

Бустингом

Постановка задачи: сделать прогноз на n дней вперед с помощью бустинга.

Особенности: надо использовать лаги (метод `transform` из `pandas`) и сделать n моделей (n =количество горизонтов)

Пример выходного датасета: Y_n – предсказанные значения

Feature_1	...	Demand_today	Demand_+1day	Demand_+2day
X1		11	12	13
X2		12	13	Y1
X3		13	Y2	Y3

Концепция

Концепция задачи состоит в том, что мы имитируем авторегрессию, при этом используем регрессионные модели. Это дает сразу несколько преимуществ относительно предсказания с помощью авторегрессии:

- Мы можем использовать более сложные модели (например бустинг)
- Мы можем исключать из обучения очень шумные данные и нам не надо городить выкалывания во временном ряде (это не очень хорошая практика, однако, когда нет другого выхода - сойдет)
- Мы можем использовать признаки в модели, которые, тяжело формализуются в рамках временного ряда
- Как следствие разнообразия моделей, можем подбирать лоссы под задачу

В основном нам необходимо сделать следующее:

- Надо выбрать гранулярность предсказания, т. е. что является меньшим неделимым в рамках модели представлением (например, в задаче спроса это может быть сложный ключ товар-магазин-день)
- Для каждого примера из выборки нужно составить набор прошлых значений по времени и набор будущих значений по времени. Примерно как в таблице ниже (зеленые ячейки необходимо предсказать).

Товар	Магазин	Дата	Спрос -2 days	Спрос -1 days	Спрос	Спрос +1 days	Спрос +2 days
X1	1	1.1.2001	9	10	11	12	13
X2	1	2.1.2001	10	11	12	13	Y1
X3	1	3.1.2001	11	12	13	Y2	Y3

- Далее строим n моделей, где n – это количество горизонтов на которые нам надо спрогнозировать (в случае таблицы выше $n=2$), и главное не забывать подавать на вход n горизонту предыдущие значения горизонтов, независимо от того фактические они или предсказанные