

Предсказание затрат

Трубицын Юрий Алексеевич

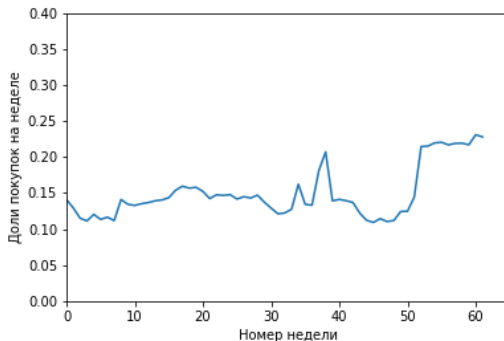
Московский Государственный Университет им. М.В. Ломоносова
факультет Вычислительной математики и кибернетики
кафедра Математических Методов Прогнозирования

Москва, 2017

Постановка задачи

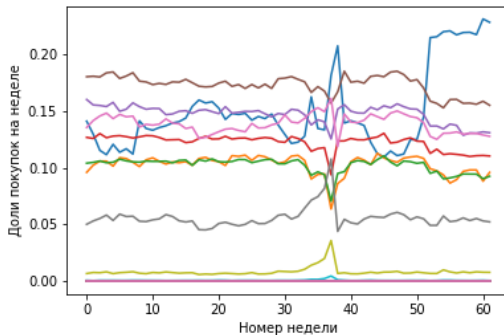
- **Дано:** данные о 110000 клиентах некоторого магазина. Данные имеют следующий формат:
 - Поле «*id*» - идентификатор клиента;
 - Поле «*date*» - дата покупки;
 - Поле «*sum*» - сумма покупки.
- **Выход:** для каждого клиента вывести сумму его первой покупки на следующей неделе. Если клиент ничего не купит в течении недели, то вернуть ноль.
- **Метрика:** ассурасу.

Наблюдения



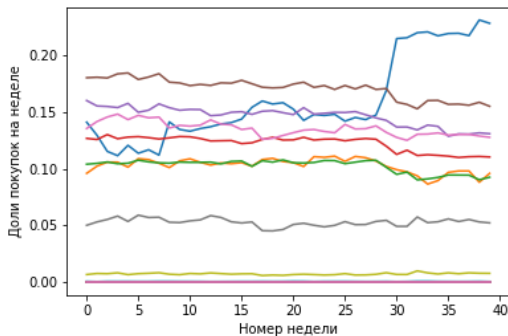
Указана доля «покупок» на сумму 0 за каждую неделю. Но это мало информативно.

Наблюдения



Указаны доли всех первых сумм покупок по неделям.
Можно «почистить» выборку.

Наблюдения



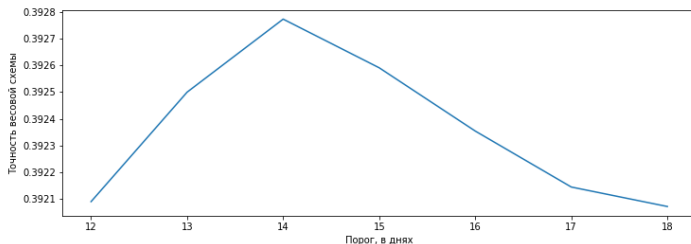
Выглядит уже более стабильно.

Что предпринималось:

- **Способ 1:** попытка спрогнозировать долю каждой суммы на следующую неделю и «расфасовать» клиентов по суммам, согласно распределению первых сумм за неделю каждого клиента;
- **Способ 2:** Использовать весовую схему с весами, равными степени номера недели.

Способ борьбы с нулевыми покупками

На графиках хорошо видно что в конце выборки многие клиенты перестали ходить в данный магазин. Способ борьбы - сумма для клиентов, которые не ходили больше некоторого порога автоматически выставляется нулем.



Способ тестирования

Способ тестирования достаточно прост

- Отсекаем от выборки последнюю неделю;
- Пытаемся предсказать по усеченной выборке;
- Сравниваем предсказание с точным ответом и выдаем долю верно предсказанных ответов.

❶ Способ 1 (восстановление долей):

- Тестирование: 0.38664
- Leaderboard: 0.39000

❷ Способ 2 (весовая схема):

- Тестирование: 0.39287
- Leaderboard: 0.39715

Что за данные?

