Отчет по заданию №1 курса AMA: Суммы покупок

Оспанов Аят

517 группа

Москва, 2016

Рассмотренные алгоритмы

Весовые схемы:

- Линейные с параметром степени:
  $w_i^0 = \left(\frac{d-i+1}{d}\right)^\delta, i \in \{1, ..., d\}, \delta \in [0; +\infty)$
- Обратно-степенные:
  $w_i^1 = \frac{1}{i^\gamma}, i \in \{1, ..., d\}, \gamma \in [0; +\infty)$

Обучение проходило по всем данным, кроме последней известной покупки. Контроль на последнем известном дне покупки.

Обработка данных

Создавался календарь для каждого пользователя и удалялись недели, на которых не было покупок. Календарь - матрица $Nx7$. Создавался календарь для дальнейшего упрощения подсчетов.

Первая попытка

Пусть D - день недели, на который нужно предсказать покупку пользователя, $s_i, i \in \{1, ..., d\}$ - покупки пользователя в D-й день i-й недели, $S_i, i \in \{1, ..., n\}$ - все покупки пользователя.
Тогда покупка пользователя на D-й день определяется по следующей схеме:

$$s = \alpha * \frac{\sum_{i=1}^{d} s_i}{d} + (1 - \alpha) * \frac{\sum_{i=1}^{n} S_i}{n}$$

Такая схема при $\alpha = 0.8$ дала MAE = 277.86249 на Public

Вторая попытка

На этот раз применялись весовые схемы и их комбинации
Покупка пользователя на D-й день определяется по следующей схеме:

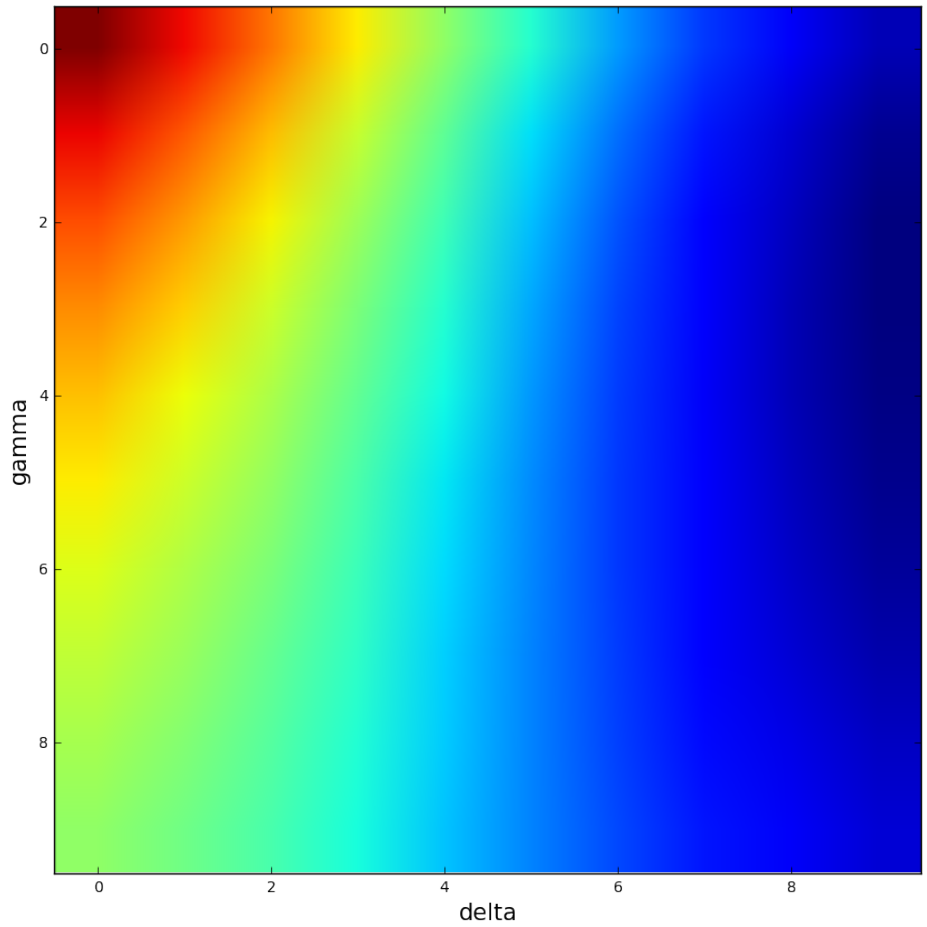
$$s = \alpha * \sum_{i=1}^{d} w_i^0 * s_i + \beta * \sum_{i=1}^{d} w_i^1 * s_i + (1 - \alpha - \beta) * \frac{\sum_{i=1}^{n} S_i}{n}$$

Настройка параметров

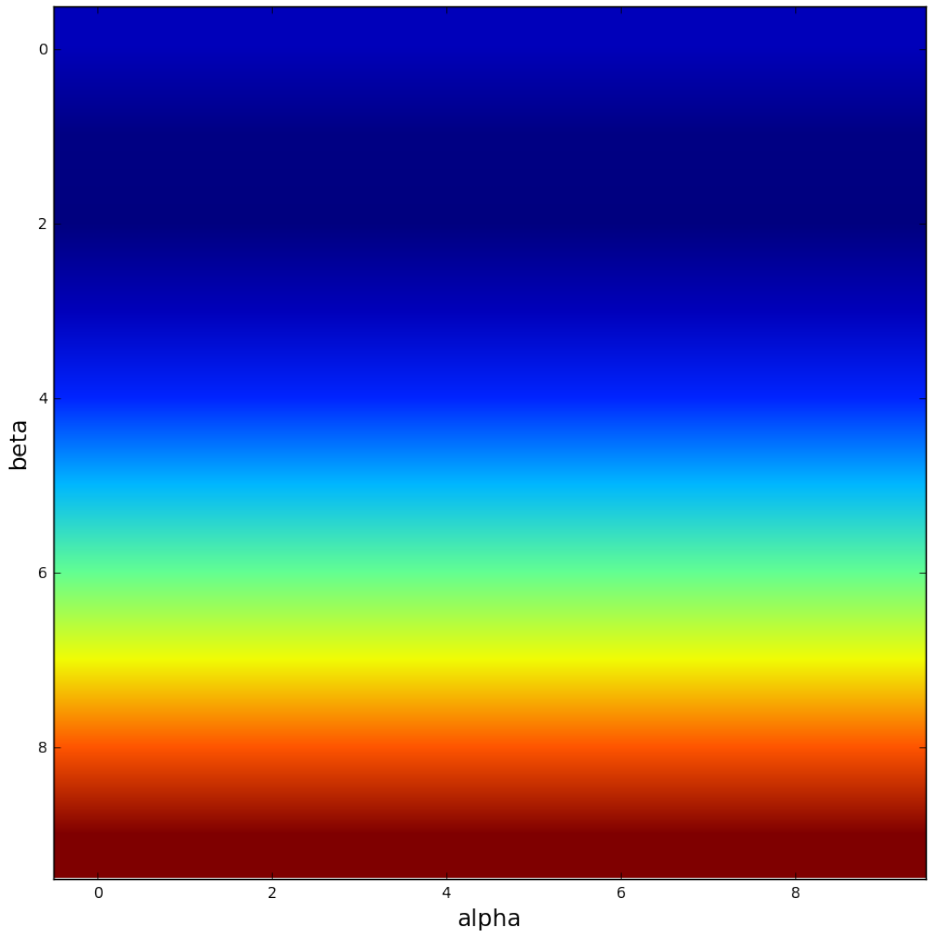
Настройка параметров велась минимизируя MAE на последнем известном дне. Побдор проходил в два этапа:

- Подбор δ, γ при фиксированных α, β
- Подбор α, β при фиксированных δ, γ

Подбор δ, γ



Подбор α, β



Результаты подбора параметров

В итоге оптимизаций получились следующие параметры:

$$\delta = 0.375, \gamma = 3.33, \alpha = 0.42, \beta = 0.33$$

$$MAE = 244.41844$$

Но это привело к тому, что мы переобучились. Public score при этих параметрах дал

$$MAE = 281.68734$$

В связи с этим, встала проблема побдора параметров. Решилось это эмпирическим путем, что привело к

$$\delta = 0.5, \gamma = 0.72, \alpha = 0.22, \beta = 0.44$$

$$MAE = 275.88369$$

Итог

Были опробованы весовые схемы предсказания суммы следующей покупки. Т.к. второй метод является более параметризируемым, удалось достигнуть неплохого результата. Но второй метод не идеален, т.к. имеет 4 гиперпараметра, что может привести к переобучению. Также выяснилось, что нужен талант и везение, чтобы настроить гиперпараметры =)

Спасибо за внимание!