

# Машинное обучение.

## Домашнее задание №17

**Задача 1.** Допустим, мы хотим обучить рекомендательную систему по истории взаимодействия  $n$  пользователей и  $m$  товаров. Рассмотрим матрицу оценок  $R \in \mathbb{R}^{n \times m}$ , в которой известны лишь некоторые элементы:  $r_{ui}$  – оценка пользователя  $1 \leq u \leq n$  для товара  $1 \leq i \leq m$ . Дополнительно предположим, что доля пар  $(u, i)$  с известными оценками относительно всех пар равна  $\alpha$ , причем эти пары равномерно распределены по матрице оценок. Рассмотрим способ предсказания неизвестных оценок на основе модели Latent Factor Model, которая использует малоранговую аппроксимацию ранга  $r$  матрицы  $R$ :

$$\tilde{r}_{ui} = p_u^T q_i, \quad p_u \in \mathbb{R}^r, \quad q_i \in \mathbb{R}^r.$$

Параметры модели можно подбирать в ходе решения задачи восстановления пропущенных значений матрицы оценок (Matrix Completion):

$$P, Q = \arg \min_{P \in \mathbb{R}^{r \times n}, Q \in \mathbb{R}^{r \times m}} \sum_{(u,i): \exists r_{ui}} (r_{ui} - p_u^T q_i)^2.$$

Рассмотрим один из наиболее известных алгоритмов решения этой оптимизационной задачи – Alternating Least Squares (ALS). Этот алгоритм поочередно фиксирует факторы  $P$  или  $Q$  и подбирает оптимальное значение другого фактора, точно решая задачу с помощью линейного метода наименьших квадратов. Например, при фиксированном  $P$ :

$$Q = \arg \min_{Q \in \mathbb{R}^{r \times m}} \sum_{(u,i): \exists r_{ui}} (r_{ui} - p_u^T q_i)^2.$$

Запишите явные формулы для вычисления оптимального  $Q$  при фиксированном  $P$ , а также покажите, что вычислительная сложность одного такого шага равна  $O(r^2 m(\alpha n + r))$ .

**Задача 2.** Рассмотрим задачу неотрицательного матричного разложения матрицы  $R \in \mathbb{R}^{n \times m}$ :

$$H, W = \arg \min_{H \in \mathbb{R}^{n \times r}, W \in \mathbb{R}^{r \times m}} \sum_{i,j} \left( R_{ij} - (HW)_{ij} \right)^2, \quad h_{ik} \geq 0, \quad w_{ik} \geq 0$$

Один из подходов к её решению – Hierarchical Alternating Least Squares (HALS). В нем по очереди оптимизируется каждый элемент матриц  $W$  и  $H$  при фиксированных

остальных элементах. Например, для  $w_{ik}$ :

$$w_{ik} = \arg \min_{W_{ik} \geq 0} \sum_{i=1}^n \sum_{j=1}^m \left( R_{ij} - (HW)_{ij} \right)^2$$

Запишите явные формулы для вычисления оптимальных значений матрицы  $W$ .

**Задача 3.** Андрей и Денис решили воспользоваться машинным обучением, чтобы улучшить продажи в своем интернет-магазине. Для этого они разработали рекомендательную систему, предлагающую пользователю три товара в момент, когда он сформировал корзину и собрался делать заказ. Эта система делает рекомендации как на основе товаров в корзине, так и на основе истории пользователя. Например, если пользователь хочет купить фотоаппарат, то система предложит ему сумку и штатив, поскольку их часто покупают в дополнение к фотоаппаратам. Или же, если данный пользователь часто покупает книги, то система может предложить ему купить вместе с фотоаппаратом недавно вышедший детектив.

Рекомендательная система у Андрея и Дениса получилась очень сложная и ресурсоемкая, и для ее стабильной работы необходимо закупить несколько мощных серверов, а также платить зарплату системному администратору, который будет следить за работой этих серверов и самой рекомсистемы. Чтобы проверить, имеет ли смысл идти на такие траты, они хотят понять, способен ли разработанный алгоритм делать правильные рекомендации.

Денис предлагает взять историю покупок пользователей, и для каждого пользователя разбить ее на две части: первые 70% взять в обучающую выборку, а последние 30% — в контроль. После этого следует настроить рекомендательный алгоритм по обучающей выборке и проверить, насколько хорошо он предсказывает пользователям покупки из контрольной выборки.

Андрей же утверждает, что нужно провести АВ-тест: разбить всех пользователей интернет-магазина на две группы и показывать рекомендации лишь в одной из групп. После этого предлагается подсчитать число купленных товаров в первой и второй группах. Если в группе, где показывались рекомендации, это число окажется больше, то рекомендательную систему следует признать полезной.

Какие преимущества и недостатки вы видите в подходах Андрея и Дениса? Кому из них следует отдать предпочтение?

**Задача 4.** В факторизационных машинах предсказание для объекта  $x \in \mathbb{R}^d$  делается по формуле

$$a(x) = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=i+1}^d x_i x_j \langle \vec{v}_i, \vec{v}_j \rangle.$$

Вычисление предсказания по этой формуле требует  $O(rd^2)$  операций, где  $r$  — размер векторов  $\vec{v}_1, \dots, \vec{v}_d$ . Покажите, что это же предсказание может быть найдено за  $O(rd)$  операций.