

## Задание 3. Рекомендательные системы

Практикум 317 группы, 2018

Начало выполнения задания: 9 апреля 2018 года.

Мягкий дедлайн: **22 апреля 2018 года, 23:59**

Жёсткий дедлайн: **6 мая 2018 года, 23:59.**

Максимальный балл: 40 (плюс бонусные баллы).

Текст задания обновлён 9 апреля 2018 г.

## Формулировка задания

Данное задание направлено на ознакомление с рекомендательными системами и конкурсами по анализу данных. Задание состоит из двух частей.

1. В первой части необходимо написать на языке Python собственную реализацию алгоритма LFM, обучаемого с помощью метода ALS, и провести все описанные ниже эксперименты. За правильное выполнение первой части можно получить 15 баллов.

Прототипы функций должны строго соответствовать прототипам, описанным в спецификации, и проходить все выданные тесты. Задание, не проходящее все выданные тесты, приравнивается к невыполненному. При написании необходимо пользоваться стандартными средствами языка Python, библиотеками `numpy`, `scipy`, `scikit-learn` и `matplotlib`, но запрещается пользоваться уже готовыми реализациями выше упомянутых методов.

2. Вторая часть задания проводится на платформе `kaggle.inclass`. Необходимо занять как можно более высокое место в конкурсе по подбору рекомендаций книг пользователям. Система выставления баллов за конкурс обговорена ниже. При решении конкурса разрешается пользоваться любыми готовыми реализациями, но запрещается использовать внешние данные. Н
3. Написать отчёт о проделанной работе (формат PDF, подготовленный в системе  $\text{\LaTeX}$ , или формат html, конвертированный из jupyter-notebook). Отчёт должен включать в себя обе части задания. Качество отчёта оценивается из 5 баллов.
4. Сдать в систему anytask .zip архив с отчётом, файл `lfm.py` с реализацией LFM и jupyter-notebook с экспериментами (если есть дополнительные модули с кодом, их тоже нужно поместить в архив). Архив должен иметь название `task3_фамилия_имя.zip`
5. Задание считается невыполненным, если не преодолен первый бейзлайн в лидерборде.

## Часть 1: Алгоритм Latent Factor Model

### Требуемые эксперименты

1. Реализуйте алгоритм Latent Factor Model с методом обучения ALS. Протестируйте алгоритм на датасете из части 2.
2. Исследуйте, как зависит качество и время работы модели от количества компонент.

## Часть 2: Рекомендация книг по истории пользователей

В этом задании вам предлагается поучаствовать в учебном конкурсе, посвящённом рекомендации книг. Доступ к контесту можно получить у преподавателя. Метрика оценивания — RMSE.

Для получения баллов за задания, необходимо преодолевать бейзлайны. Всего в конкурсе 4 бейзлайна, каждый бейзлайн оценивается в 5 баллов. При сдаче работы, необходимо описать в отчёте:

- выбранную методику тестирования
- найденные в данных закономерности
- краткое описание выбранных моделей/использованных библиотек

- качество моделей на локальном контроле и на public/privat лидерборде

После окончания конкурса трём лучшим (на private лидерборде) из числа преодолевших **все** бейзлайны назначается 15, 13 и 11 баллов соответственно. Остальным участникам, преодолевшим все бейзлайны, назначается от 1 до 10 бонусных баллов. Количество баллов линейно зависит от позиции в лидерборде. Участникам, проявившим какую-либо полезную активность (выкладывание решений в открытый доступ, активное обсуждение в форуме соревнования), могут быть назначены дополнительные бонусные баллы (на усмотрение преподавателя).