

## Попытка №1 (3-v1.ipynb)

Первая идея была использовать RandomForestClassifier алгоритм.

Данные, на котором обучалось:

- Матрица: [все фильмы] x [жанр, есть ли признак 1, есть ли признак 2, ..., есть ли признак n]

Результат, на котором обучалось:

- Вектор: [есть ли лайк к фильму 1, есть ли лайк к фильму 2, ... , есть ли лайк к фильму n]

Проблемы, из-за которых это решение не работает:

- Неправильный физический смысл результата: получилось, что отсутствие лайка означает, что пользователь никогда не лайкнет (= дислайк), а должно означать, что он пока ещё не лайкнул.
- Приходится грубо обрезать данные (выбирать k случайных фильмов), иначе обучение будет длиться вечно.

## Попытка №2 (3-v2.ipynb)

Используются два способа для вычисления вероятности, и в итоге выбирается максимальное из рассчитанных вероятностей лайка.

Для начала создаем 3 матрички:

1. фильмы x (жанры + признаки)
2. люди x фильмы
3. люди x (жанры + признаки)

### Первый способ: по другим пользователям

Применяем SVD к матрице (люди x фильмы), используя 100 компонент. (почему? см. материал 3 в списке внизу, слайд 65, а также материал 5) Компоненты обозначают свойства элементов исходной матрицы. Получаем три матрицы: первая U (люди x компоненты), вторая  $\sigma$  (компоненты x компоненты) и третья V транспонированная (компоненты x фильмы).

Чтобы узнать вероятность, узнаем косинус угла между векторами ( U[человек, все компоненты] \*  $\sigma$  ) и V[все компоненты, фильм] (почему? см. материал 11). Узнаем косинус из скалярного произведения:  $\langle \mathbf{a}, \mathbf{b} \rangle = |\mathbf{a}| \cdot |\mathbf{b}| \cdot \cos \angle(\mathbf{a}, \mathbf{b})$ . Значение модуля косинуса угла между 0 и 1, так что его можно использовать для вероятности. Вероятность = косинус угла. Например, если вектора совпадают, то есть угол = 0, это означает, что предсказание правильное.

## Второй способ: по признакам фильма

Берем соответствующие строчки из матриц (фильмы  $\times$  (жанры + признаки)) и (люди  $\times$  (жанры + признаки)). Как и в первом способе находим косинус угла между этими вектор-строками и получаем вероятность.

## Список использованных материалов

1. Введение в машинное обучение  
<https://www.coursera.org/learn/vvedenie-mashinnoe-obuchenie>
2. Classifier comparison  
[http://scikit-learn.org/stable/auto\\_examples/classification/plot\\_classifier\\_comparison.html](http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html)
3. Recommender Systems (Machine Learning Summer School 2014 @ CMU)  
<http://www.slideshare.net/xamat/recommender-systems-machine-learning-summer-school-2014-cmu>
4. Singular value decomposition  
[https://en.wikipedia.org/wiki/Singular\\_value\\_decomposition](https://en.wikipedia.org/wiki/Singular_value_decomposition)
5. Городской АД: школьники и студенты  
<https://habrahabr.ru/company/goto/blog/305526/>
6. Lecture: The Singular Value Decomposition (SVD)  
<https://www.youtube.com/watch?v=EokL7E6o1AE>
7. Computing the Singular Value Decomposition | MIT 18.06SC Linear Algebra, Fall 2011  
<https://www.youtube.com/watch?v=cOUTpqlX-Xs>
8. Ортогональные матрицы  
[http://www.slideshare.net/tm\\_ssau/ss-25992336](http://www.slideshare.net/tm_ssau/ss-25992336)
9. importance of PCA or SVD in machine learning  
<http://stackoverflow.com/questions/9590114/importance-of-pca-or-svd-in-machine-learning>
10. What is an intuitive explanation of singular value decomposition (SVD)?  
<https://www.quora.com/What-is-an-intuitive-explanation-of-singular-value-decomposition-SVD/answer/Jason-Liu-21?srid=xmCp>
11. How do I use the SVD in collaborative filtering?  
<http://stats.stackexchange.com/questions/31096/how-do-i-use-the-svd-in-collaborative-filtering/31101#31101>
12. Restricted Boltzmann Machines (RBM)  
<http://deeplearning.net/tutorial/rbm.html#contrastive-divergence-cd-k>
13. Introduction to Restricted Boltzmann Machines  
<http://blog.echen.me/2011/07/18/introduction-to-restricted-boltzmann-machines>
14. Реализация Restricted Boltzmann machine на c#  
<https://habrahabr.ru/post/159909/>

## 15. Метрики ранжирования

[https://stackedit.io/viewer#!provider=gist&gistId=de556eedb4dee5d7ca36eb922be40217&filename=ranking\\_metrics](https://stackedit.io/viewer#!provider=gist&gistId=de556eedb4dee5d7ca36eb922be40217&filename=ranking_metrics)