

# PCML CS-433: Recommender System

Gael Lederrey, SCIPER ???, gael.lederrey@epfl.ch  
Stefano Savarè, SCIPER ???, stefano.savare@epfl.ch  
Joachim Muth, SCIPER 214757, joachim.muth@epfl.ch

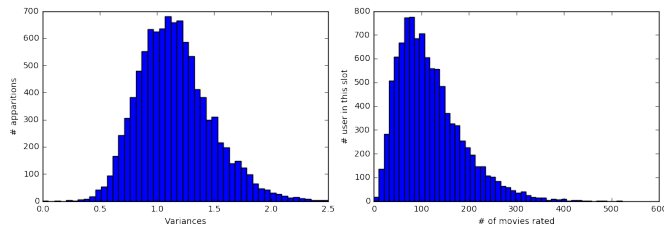
*School of Computer and Communication Sciences, EPF Lausanne, Switzerland*

## Abstract—

### I. DATA DESCRIPTION

The data represent ratings from 10'000 users on 1'000 movies in an integer scale from 1 to 5. This scale represent the number of *stars* given by the users, 1 being the lowest grade and 5 the best.

The training set used to train our algorithm contains 1'176'952 ratings which represent around 12% of possible filled ratings. An other 1'176'952 ratings are hidden from us and must be predicted by our recommender algorithm.



(a) Distribution of variances of ratings per user. (b) Number of movies rated per user.

Figure 1: Statistical description of data

### II. DATA PREPROCESSING

### III. MODEL SELECTION

### IV. RESULT

### V. DISCUSSION