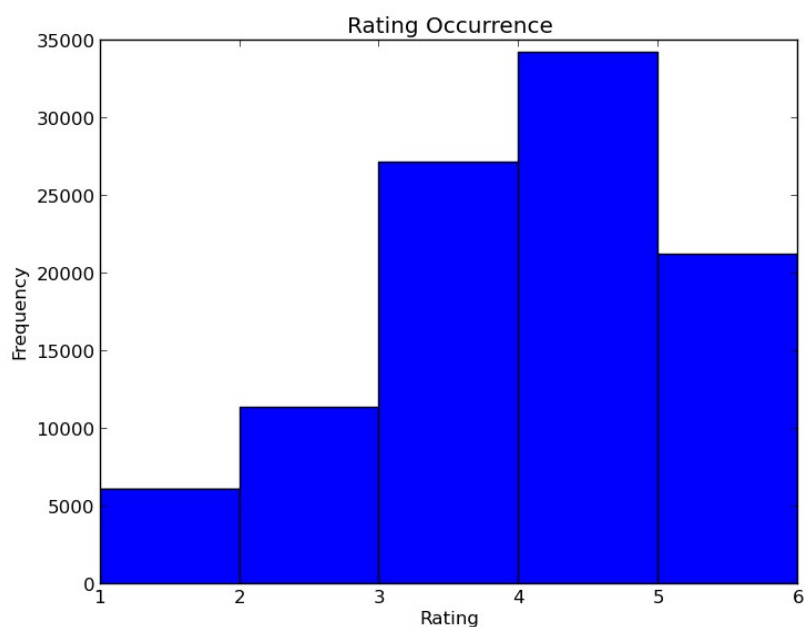Transparent Movie Recommendation System
Weekly Report
8 September 2015
Neil Getty
Mustafa Bilgic

Code changes this week:
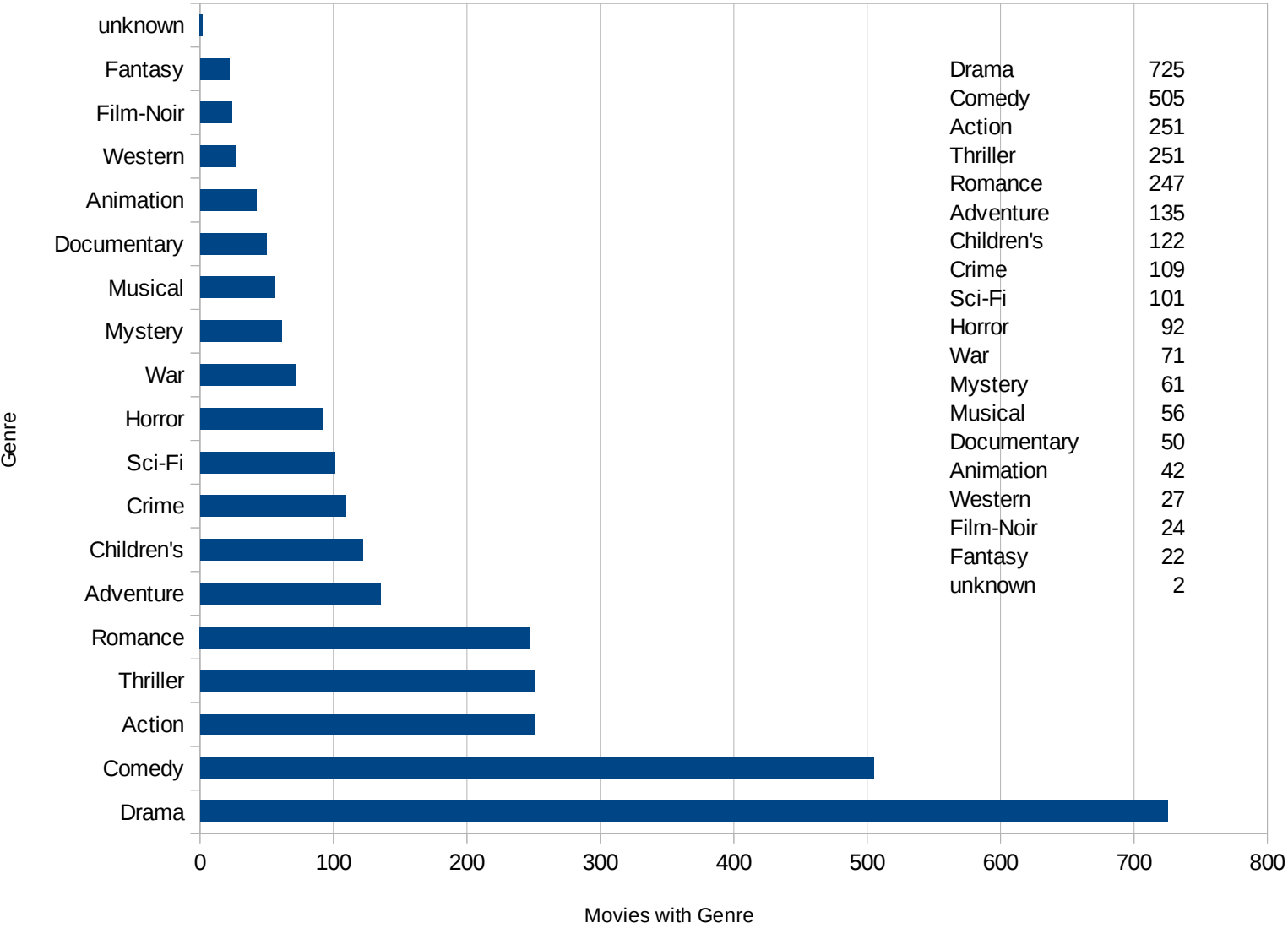
Average rating is now scaled using sklearn.preprocessing.

Below are charts and tables depicting data features:

This chart depicts the ratings of **943** users.
The rating frequency distribution is left-skewed.
It is likely people tend to review movies they enjoy.

Genre Occurance

Drama	725
Comedy	505
Action	251
Thriller	251
Romance	247
Adventure	135
Children's	122
Crime	109
Sci-Fi	101
Horror	92
War	71
Mystery	61
Musical	56
Documentary	50
Animation	42
Western	27
Film-Noir	24
Fantasy	22
unknown	2

Genre

Movies with Genre

Drama and comedy are by far the most frequent genres of the *19*.

Top *10* least/most frequent keywords of *370*

45	paris-france
45	suicide-attempt
45	greed
45	ship
45	newspaper-headline
45	basketball
45	product-placement
45	semiautomatic-pistol
45	champagne
45	1940s
270	friendship
274	title-spoken-by-character
276	cult-film
287	violence
292	love
296	based-on-novel
300	husband-wife-relationship
320	death
414	murder
480	independent-film

Below are charts and tables depicting model results:

Models were trained with cross-validation using **5** folds

MAE given *alphas* in *range(1,20)*.

Alpha	MAE
1	0.8793379821
2	0.8625366464
3	0.8435265508
4	0.8322018457
5	0.8265751307
6	0.8301550595
7	0.8139420623
8	0.8169360833
9	0.817121407
10	0.8106828301
11	0.8195926043
12	0.8191601729
13	0.8064503482
14	0.8150722365
15	0.8082146932
16	0.8111721182
17	0.8109946119
18	0.8150954161
19	0.8165423226

RidgeCv
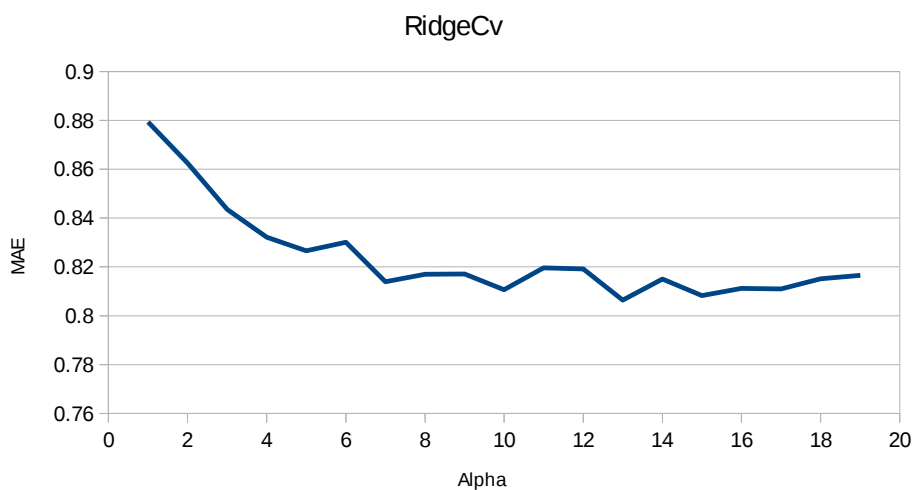
Alpha	MAE
0.01	0.8974925108
0.02	0.8392563215
0.03	0.8130726458
0.04	0.8074623706
0.05	0.7941261782
0.06	0.7965255467
0.07	0.7943692814
0.08	0.786841095
0.09	0.7889524035
0.1	0.784578596
0.11	0.7945739917
0.12	0.794822421
0.13	0.7962395991
0.14	0.7918176769
0.15	0.8104923694
0.16	0.794568814
0.17	0.8050769032
0.18	0.8078366815
0.19	0.8079446051
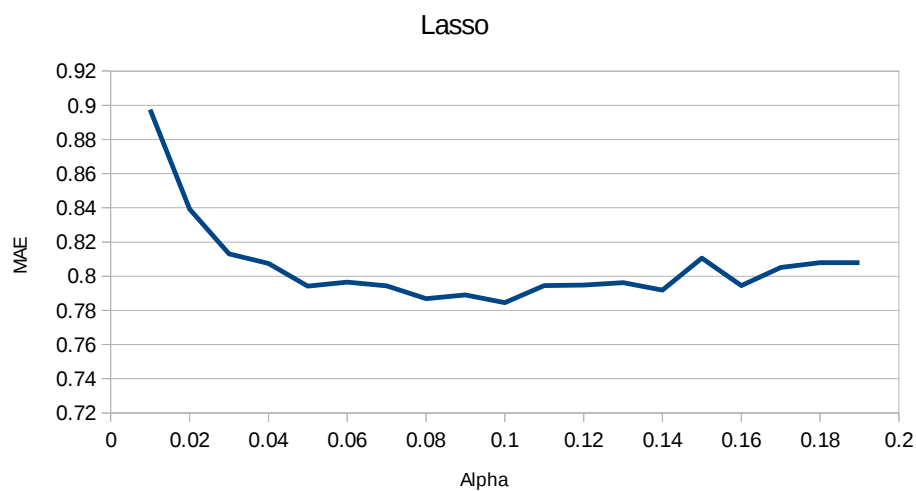
Lasso

Alphas that minimize MAE: *15* for Ridge regression and *0.1* for Lasso
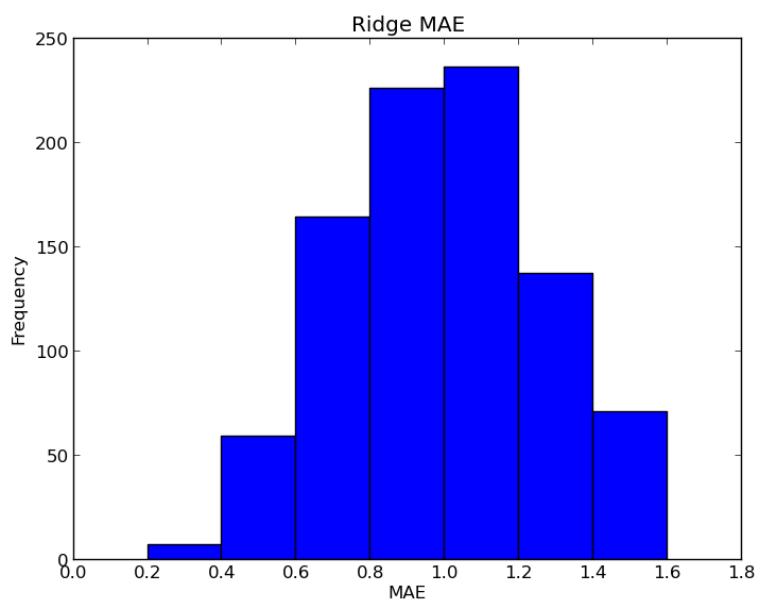
MAE frequency for Ridge and Lasso regression.
Average MAE approximately *0.80* for Ridge and *0.78* for Lasso
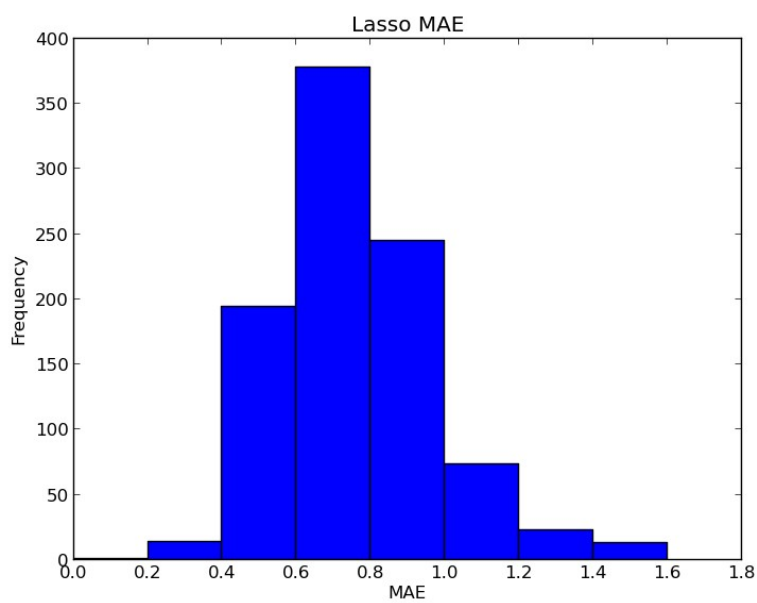
The top ten lowest/highest MAE for LassoCV with alpha = .01 and cv = 5.

849	0.2283653846
33	0.3018189997
507	0.3174908855
118	0.3393755353
552	0.3442255691
644	0.3469387755
555	0.3604217709
594	0.3634998311
923	0.3683163344
25	0.370447101
86	1.5058984012
61	1.5551890161
471	1.6036284058
166	1.6142890415
36	1.6673940731
739	1.7064231571
50	1.8041269897
762	1.8692021111
229	1.9643266731
732	2.0999669787
127	2.10625