

Neil Getty
Mustafa Bilgic
Weekly Report
13 October 2015

Corrected cross validation.

Below are tables showing the top/bottom 10 model MAE per user, 5 *fold* cross validation was used.

Ridge Regression

| User ID | MAE |
|-----------|--------|
| 849 | 0.2491 |
| 355 | 0.3165 |
| 688 | 0.3328 |
| 644 | 0.3439 |
| 384 | 0.3707 |
| 552 | 0.3708 |
| 628 | 0.4314 |
| 10 | 0.4337 |
| 800 | 0.4343 |
| 507 | 0.4433 |
| 219 | 1.3074 |
| 166 | 1.3076 |
| 448 | 1.3089 |
| 562 | 1.3120 |
| 589 | 1.3499 |
| 739 | 1.3689 |
| 179 | 1.3931 |
| 762 | 1.4277 |
| 229 | 1.4514 |
| 38 | 1.4757 |
| Average: | 0.7711 |
| Variance: | 0.0341 |

Lasso Regression

| User ID | MAE |
|-----------|--------|
| 849 | 0.2535 |
| 355 | 0.2809 |
| 644 | 0.2878 |
| 688 | 0.3368 |
| 552 | 0.3611 |
| 25 | 0.3962 |
| 507 | 0.4032 |
| 118 | 0.4106 |
| 384 | 0.4156 |
| 275 | 0.4202 |
| 732 | 1.3503 |
| 562 | 1.3652 |
| 475 | 1.3883 |
| 589 | 1.4207 |
| 38 | 1.4309 |
| 179 | 1.4533 |
| 50 | 1.4699 |
| 228 | 1.4884 |
| 471 | 1.5152 |
| 229 | 1.5555 |
| Average: | 0.7671 |
| Variance: | 0.0378 |

Models used for tables below trained using *ridge regression* and *5 fold* cross validation.

Each user result table only displays the top/lowest five movies by model error.

User 849 Results (Low MAE) Total Movies = 23

| Movie ID | Average Rating | Error from Avg | User Rating | Model Prediction | Model Error |
|-----------|----------------|----------------|-------------|------------------|-------------|
| 174 | 4.2524 | 0.7476 | 5.0000 | 4.9992 | 0.0008 |
| 133 | 3.8713 | 1.1287 | 5.0000 | 4.9909 | 0.0091 |
| 172 | 4.2044 | 0.7956 | 5.0000 | 5.0374 | 0.0374 |
| 625 | 3.3293 | 1.6707 | 5.0000 | 4.9558 | 0.0442 |
| 421 | 3.5943 | 1.4057 | 5.0000 | 4.9167 | 0.0833 |
| 27 | 3.1053 | 1.8947 | 5.0000 | 4.6985 | 0.3015 |
| 38 | 3.0083 | 1.9917 | 5.0000 | 4.6753 | 0.3247 |
| 568 | 3.6478 | 0.3522 | 4.0000 | 4.9457 | 0.9457 |
| 427 | 4.2922 | 0.2922 | 4.0000 | 5.0565 | 1.0565 |
| 406 | 2.4490 | 1.5510 | 4.0000 | 5.0622 | 1.0622 |
| Average: | 3.6157 | 1.2792 | 4.8696 | 4.8850 | 0.2543 |
| Variance: | 0.1958 | 0.2050 | 0.1134 | 0.0110 | 0.0956 |

The following tables were compiled by iterating through the weights for each movie (where prediction error fell below average) and storing the largest positive/negative features.

Most Positive Weights

| Feature | Times Largest Weight |
|------------------------------|----------------------|
| explosion | 3 |
| mother-daughter-relationship | 2 |
| Average Rating | 2 |
| based-on-novel | 2 |
| Romance | 2 |
| divorce | 1 |
| Action | 1 |
| family-relationships | 1 |
| blockbuster | 1 |
| brother-brother-relationship | 1 |
| Animation | 1 |

Most Negative Weights

| Feature | Times Lowest Weight |
|----------------------|---------------------|
| cult-film | 5 |
| Average Rating | 3 |
| Sci-Fi | 3 |
| sex | 2 |
| black-comedy | 2 |
| dysfunctional-family | 1 |

The following tables show the non-zero features for the movies with the highest and lowest prediction error.

Movie 174 (Error = .0008)

| Feature | Weights |
|---|---------|
| explosion | 0.0189 |
| Average Rating | 0.0189 |
| blockbuster | 0.0155 |
| crying | 0.0112 |
| Adventure | 0.0092 |
| Action | 0.0087 |
| children | 0.0071 |
| bus | 0.0063 |
| train | 0.0061 |
| new-york | 0.0057 |
| fear | 0.0055 |
| machine-gun | 0.0052 |
| reporter | 0.0046 |
| camera | 0.0046 |
| three-word-title | 0.0039 |
| escape | 0.0036 |
| f-word | 0.0035 |
| wheelchair | 0.0035 |
| character-repeating-someone-else's-dialogue | 0.0035 |
| Fishburne, Laurence | 0.0031 |
| disarming-someone | 0.0031 |
| extramarital-affair | 0.0027 |
| unfaithfulness | 0.0027 |
| surrealism | 0.0017 |
| falling-from-height | 0.0017 |
| liar | 0.0009 |
| christmas | 0.0009 |
| birthday | 0.0009 |
| childbirth | 0.0009 |
| jail | 0.0009 |
| food | 0.0009 |
| father-daughter-relationship | -0.0017 |
| photograph | -0.0109 |
| betrayal | -0.0117 |
| urination | -0.0164 |
| snow | -0.0164 |
| shot-in-the-forehead | -0.0221 |
| femme-fatale | -0.0259 |
| funeral | -0.0263 |
| black-comedy | -0.0411 |
| cult-film | -0.0426 |

Movie 406 (Error = 1.0622)

| Feature | Weights |
|----------------|---------|
| based-on-novel | 0.0174 |
| blockbuster | 0.0155 |
| female-nudity | 0.0101 |
| gunfight | 0.0083 |
| teacher | 0.0075 |
| fire | 0.0071 |
| convertible | 0.0063 |
| Horror | 0.0052 |
| camera | 0.0046 |
| Thriller | 0.0040 |
| 19th-century | 0.0009 |
| baby | 0.0009 |
| photograph | -0.0109 |
| actress | -0.0263 |
| Average Rating | -0.0347 |

User 737 (Average MAE) Total Movies = 33

| Movie ID | Average Rating | Error from Avg | User Rating | Model Prediction | Model Error |
|-----------|----------------|----------------|-------------|------------------|-------------|
| 169 | 4.4661 | 0.4661 | 4.0000 | 3.9933 | 0.0067 |
| 475 | 3.8840 | 0.1160 | 4.0000 | 4.0368 | 0.0368 |
| 100 | 4.1555 | 0.8445 | 5.0000 | 4.9376 | 0.0624 |
| 64 | 4.4452 | 0.4452 | 4.0000 | 4.0860 | 0.0860 |
| 160 | 3.4638 | 0.5362 | 4.0000 | 3.8182 | 0.1818 |
| 258 | 3.8035 | 1.1965 | 5.0000 | 3.6965 | 1.3035 |
| 186 | 3.8367 | 1.1633 | 5.0000 | 3.4957 | 1.5043 |
| 174 | 4.2524 | 2.2524 | 2.0000 | 3.7848 | 1.7848 |
| 96 | 4.0068 | 2.0068 | 2.0000 | 4.2006 | 2.2006 |
| 501 | 3.4959 | 2.4959 | 1.0000 | 3.7400 | 2.7400 |
| Average: | 3.9921 | 0.7581 | 3.9697 | 3.9183 | 0.7751 |
| Variance: | 0.0708 | 0.3561 | 0.9991 | 0.1138 | 0.3775 |

Most Positive Weights

| Feature | Times Largest Weight |
|------------------------------|----------------------|
| brother-brother-relationship | 6 |
| restaurant | 2 |
| independent-film | 2 |
| liar | 1 |
| violence | 1 |
| blackmail | 1 |
| 1950s | 1 |
| surrealism | 1 |
| adultery | 1 |
| party | 1 |
| Average Rating | 1 |

Most Negative Weights

| Feature | Times Lowest Weight |
|---------------------------|---------------------|
| Average Rating | 4 |
| title-spoken-by-character | 3 |
| trial | 2 |
| kidnapping | 2 |
| coffin | 1 |
| good-versus-evil | 1 |
| Adventure | 1 |
| 1970s | 1 |
| ship | 1 |
| Animation | 1 |
| DeVito, Danny | 1 |
| 1930s | 1 |

Movie 169 (Error = 0.0067)

| Feature | Weights |
|------------------------------|---------|
| brother-brother-relationship | 0.0918 |
| Average Rating | 0.0860 |
| betrayal | 0.0298 |
| christmas | 0.0285 |
| singer | 0.0260 |
| child-in-peril | 0.0217 |
| redemption | 0.0166 |
| police-officer | 0.0160 |
| dysfunctional-family | 0.0127 |
| crying | 0.0109 |
| police-officer-killed | 0.0102 |
| warrior | 0.0069 |
| based-on-play | 0.0021 |
| conspiracy | 0.0012 |
| boy | 0.0003 |
| Comedy | -0.0047 |
| action-hero | -0.0180 |
| teacher | -0.0198 |
| torture | -0.0428 |
| Animation | -0.0778 |

Movie 501 (Error = 2.7400)

| Feature | Weights |
|----------------|---------|
| Children's | 0.0456 |
| liar | 0.0442 |
| gay-slur | 0.0360 |
| betrayal | 0.0298 |
| rooftop | 0.0209 |
| paranoia | 0.0114 |
| orphan | 0.0107 |
| based-on-play | -0.0022 |
| camera | -0.0032 |
| fire | -0.0295 |
| kiss | -0.0310 |
| ship | -0.0460 |
| Animation | -0.0778 |
| Average Rating | -0.0900 |

| User 38 (High MAE) Total Movies = 121 | | | | | | |
|--|----------------|----------------|-------------|------------------|-------------|--|
| Movie ID | Average Rating | Error from Avg | User Rating | Model Prediction | Model Error | |
| 418 | 3.5814 | 1.4186 | 5.0000 | 4.9952 | 0.0048 | |
| 28 | 3.9312 | 0.0688 | 4.0000 | 4.0417 | 0.0417 | |
| 1032 | 2.8750 | 1.1250 | 4.0000 | 3.8683 | 0.1317 | |
| 404 | 3.6733 | 1.3267 | 5.0000 | 4.8336 | 0.1664 | |
| 71 | 3.7818 | 1.2182 | 5.0000 | 4.7992 | 0.2008 | |
| 257 | 3.7459 | 2.7459 | 1.0000 | 4.3263 | 3.3263 | |
| 432 | 3.7701 | 2.7701 | 1.0000 | 4.5419 | 3.5419 | |
| 526 | 3.8306 | 2.8306 | 1.0000 | 4.9350 | 3.9350 | |
| 403 | 3.4279 | 2.4279 | 1.0000 | 4.9650 | 3.9650 | |
| 195 | 3.9336 | 2.9336 | 1.0000 | 5.1148 | 4.1148 | |
| Average: | 3.1901 | 1.5144 | 3.7190 | 3.6844 | 1.4748 | |
| Variance: | 0.3711 | 0.4529 | 2.5326 | 0.4260 | 0.7780 | |

Most Positive Weights

| Feature | Times Largest Weight |
|-----------------------|----------------------|
| Average Rating | 12 |
| Action | 11 |
| Animation | 9 |
| boy | 4 |
| 1980s | 3 |
| blockbuster | 3 |
| Horror | 2 |
| female-frontal-nudity | 2 |
| child's-point-of-view | 2 |
| teenage-boy | 2 |
| Adventure | 2 |
| Ventimiglia, John (I) | 1 |
| 1990s | 1 |
| Western | 1 |
| motorcycle | 1 |
| teenager | 1 |
| Mystery | 1 |
| electrocution | 1 |
| independent-film | 1 |
| surrealism | 1 |
| orphan | 1 |
| kitchen | 1 |
| bridge | 1 |
| disguise | 1 |
| dog | 1 |
| conspiracy | 1 |
| child-in-peril | 1 |
| swimming | 1 |

Most Negative Weights

| Feature | Times Lowest Weight |
|--------------------------|---------------------|
| Average Rating | 11 |
| Romance | 11 |
| Thriller | 10 |
| Children's | 9 |
| cult-film | 7 |
| three-word-title | 4 |
| television | 3 |
| drunkenness | 3 |
| knife | 2 |
| 1950s | 2 |
| surprise-ending | 2 |
| shot-in-the-head | 1 |
| title-directed-by-female | 1 |
| car-accident | 1 |
| Comedy | 1 |
| high-school | 1 |
| police-officer | 1 |

Movie 418 (Error = 0.0048)

| Feature | Weights |
|--------------------------|---------|
| Animation | 0.2317 |
| 1990s | 0.1224 |
| Average Rating | 0.1040 |
| motorcycle | 0.0931 |
| conspiracy | 0.0722 |
| two-word-title | 0.0656 |
| train | 0.0574 |
| class-differences | 0.0549 |
| unfaithfulness | 0.0458 |
| Baldwin, Alec | 0.0455 |
| sex | 0.0432 |
| horse | 0.0338 |
| photographer | 0.0314 |
| based-on-play | 0.0307 |
| anti-hero | 0.0208 |
| black-comedy | 0.0159 |
| ship | 0.0035 |
| warrior | -0.0019 |
| combat | -0.0053 |
| face-slap | -0.0074 |
| ambush | -0.0112 |
| Musical | -0.0142 |
| california | -0.0316 |
| Hedaya, Dan | -0.0389 |
| blackmail | -0.0396 |
| tragedy | -0.0751 |
| explosion | -0.1101 |
| shot-in-the-head | -0.1178 |
| satire | -0.1347 |
| police | -0.1417 |
| Children's | -0.1511 |
| title-directed-by-female | -0.2140 |

Movie 501 (Error = 4.1148)*Truncated from 84 features to 30*

| Feature | Weights |
|---------------------------|---------|
| Average Rating | 0.1977 |
| Action | 0.1908 |
| husband-wife-relationship | 0.1901 |
| female-frontal-nudity | 0.1632 |
| seduction | 0.1270 |
| surrealism | 0.1233 |
| blockbuster | 0.1230 |
| rescue | 0.1085 |
| scientist | 0.0968 |
| interrogation | 0.0918 |
| swimming | 0.0905 |
| conspiracy | 0.0722 |
| chase | 0.0659 |
| electrocution | 0.0593 |
| greed | 0.0591 |
| wealth | -0.0558 |
| pay-phone | -0.0567 |
| 1970s | -0.0640 |
| wheelchair | -0.0649 |
| evil-man | -0.0656 |
| convertible | -0.0723 |
| cemetery | -0.0725 |
| underwear | -0.0984 |
| explosion | -0.1101 |
| hallucination | -0.1511 |
| knife | -0.1558 |
| three-word-title | -0.1782 |
| surprise-ending | -0.1804 |
| car-accident | -0.2035 |
| Thriller | -0.2230 |
| television | -0.2384 |

Models used for tables below trained using *lasso regression* and *5 fold* cross validation.

Each user result table only displays the top/lowest five movies by model error.

User 849 Results (Low MAE) Total Movies = 23

| Movie ID | Average Rating | Error from Avg | User Rating | Model Prediction | Model Error |
|-----------|----------------|----------------|-------------|------------------|-------------|
| 588 | 3.7921 | 1.2079 | 5.0000 | 4.9630 | 0.0370 |
| 421 | 3.5943 | 1.4057 | 5.0000 | 4.9491 | 0.0509 |
| 234 | 3.7750 | 1.2250 | 5.0000 | 4.8889 | 0.1111 |
| 298 | 3.7680 | 1.2320 | 5.0000 | 4.8889 | 0.1111 |
| 288 | 3.4414 | 1.5586 | 5.0000 | 4.8889 | 0.1111 |
| 197 | 4.1046 | 0.8954 | 5.0000 | 4.8333 | 0.1667 |
| 15 | 3.7782 | 1.2218 | 5.0000 | 4.8333 | 0.1667 |
| 406 | 2.4490 | 1.5510 | 4.0000 | 4.8889 | 0.8889 |
| 568 | 3.6478 | 0.3522 | 4.0000 | 4.9528 | 0.9528 |
| 427 | 4.2922 | 0.2922 | 4.0000 | 4.9983 | 0.9983 |
| Average: | 3.6157 | 1.2792 | 4.8696 | 4.8700 | 0.2466 |
| Variance: | 0.2047 | 0.2143 | 0.1186 | 0.0026 | 0.0785 |

No nonzero features

User 737 (Average MAE) Total Movies = 33

| Movie ID | Average Rating | Error from Avg | User Rating | Model Prediction | Model Error |
|-----------|----------------|----------------|-------------|------------------|-------------|
| 173 | 4.1728 | 0.1728 | 4.0000 | 4.0222 | 0.0222 |
| 428 | 3.8182 | 0.1818 | 4.0000 | 3.9737 | 0.0263 |
| 475 | 3.8840 | 0.1160 | 4.0000 | 4.0678 | 0.0678 |
| 64 | 4.4452 | 0.4452 | 4.0000 | 4.1287 | 0.1287 |
| 32 | 3.7901 | 0.2099 | 4.0000 | 3.8277 | 0.1723 |
| 186 | 3.8367 | 1.1633 | 5.0000 | 3.6153 | 1.3847 |
| 258 | 3.8035 | 1.1965 | 5.0000 | 3.6010 | 1.3990 |
| 174 | 4.2524 | 2.2524 | 2.0000 | 4.0317 | 2.0317 |
| 96 | 4.0068 | 2.0068 | 2.0000 | 4.2592 | 2.2592 |
| 501 | 3.4959 | 2.4959 | 1.0000 | 3.9737 | 2.9737 |
| Average: | 3.9921 | 0.7581 | 3.9697 | 3.9325 | 0.8083 |
| Variance: | 0.0731 | 0.3673 | 1.0303 | 0.0682 | 0.4551 |

Most Positive Weights

| Feature | Times Largest Weight |
|------------------------------|----------------------|
| brother-brother-relationship | 6 |
| black-comedy | 3 |
| based-on-novel | 1 |
| father-daughter-relationship | 1 |

Most Negative Weights

| Feature | Times Lowest Weight |
|------------------|---------------------|
| good-versus-evil | 3 |
| action-hero | 1 |

Movie 173 (Error = 0.0222)

| Feature | Weights |
|------------------------------|---------|
| father-daughter-relationship | 0.0642 |
| restaurant | 0.0069 |

Movie 501 (Error = 2.9737)

No nonzero features

User 38 (High MAE) Total Movies = 121

| Movie ID | Average Rating | Error from Avg | User Rating | Model Prediction | Model Error |
|-----------|----------------|----------------|-------------|------------------|-------------|
| 28 | 3.9312 | 0.0688 | 4.0000 | 3.9652 | 0.0348 |
| 672 | 2.7231 | 0.2769 | 3.0000 | 3.1347 | 0.1347 |
| 424 | 1.3158 | 1.6842 | 3.0000 | 3.1729 | 0.1729 |
| 411 | 2.9141 | 0.0859 | 3.0000 | 3.2264 | 0.2264 |
| 1032 | 2.8750 | 1.1250 | 4.0000 | 3.7365 | 0.2635 |
| 257 | 3.7459 | 2.7459 | 1.0000 | 4.0796 | 3.0796 |
| 211 | 3.9126 | 2.9126 | 1.0000 | 4.1908 | 3.1908 |
| 195 | 3.9336 | 2.9336 | 1.0000 | 4.2047 | 3.2047 |
| 526 | 3.8306 | 2.8306 | 1.0000 | 4.2079 | 3.2079 |
| 432 | 3.7701 | 2.7701 | 1.0000 | 4.2928 | 3.2928 |
| Average: | 3.1901 | 1.5144 | 3.7190 | 3.7335 | 1.4294 |
| Variance: | 0.3742 | 0.4567 | 2.5537 | 0.1208 | 0.5399 |

Most Positive Weights

| Feature | Times Largest Weight |
|---------------------------|----------------------|
| Animation | 9 |
| rescue | 8 |
| blockbuster | 6 |
| Western | 1 |
| husband-wife-relationship | 1 |

Most Negative Weights

| Feature | Times Lowest Weight |
|------------|---------------------|
| Romance | 13 |
| police | 4 |
| psychopath | 1 |

Movie 28 (Error = 0.0348)

| Feature | Weights |
|---------|---------|
| rescue | 0.1663 |

Movie 432 (Error = 3.2928)

| Feature | Weights |
|-----------|--------------|
| Animation | 0.3338304332 |