

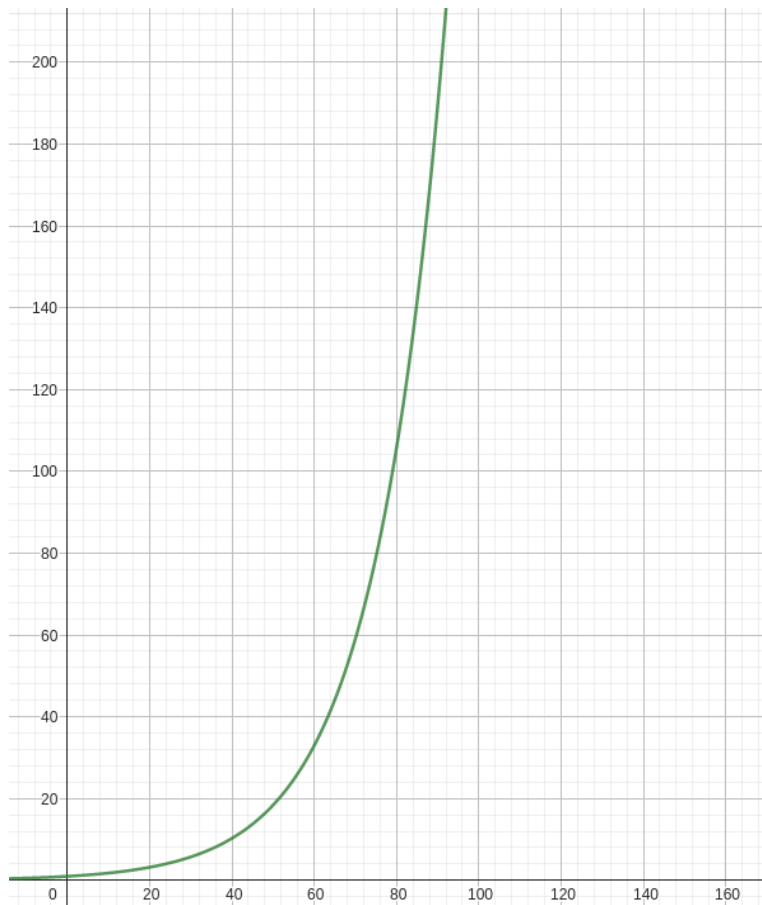
ΟΝΟΜΑΤΕΠΩΝΥΜΟ: ΖΕΡΒΑΣ ΜΙΧΑΗΛ  
ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: ics20015

### HW5

Η καλύτερη επίδοση του συστήματος για τις συγκεκριμένες μετρικές βρέθηκε να είναι για  $k=15$  και  $\text{train}=90\%$ . Για αυτόν τον συνδυασμό  $k$  έχουμε τη μεγαλύτερη τιμή Precision με ταυτόχρονη υψηλή τιμή (συγκριτικά με τις άλλες μετρήσεις) Recall καθώς και το μικρότερο δυνατό MAE. Γενικά οι αποκλίσεις μεταξύ των μετρικών, για όλες τις επιλογές που το έτρεξα, (όπως φαίνεται και στη συνέχεια) είναι πολύ μικρές. Επίσης παρατηρούμε ότι το MAE αυξάνεται για μεγάλες τιμές του  $k$  όπως επίσης και το Recall, ενώ το Precision μειώνεται. Για τα διάφορα ποσοστά του train παρατηρούμε ότι, όσο ανεβαίνει το ποσοστό βελτιώνονται και οι μετρικές.

Όσον αφορά τις επιμέρους λεπτομέρειες της υλοποίησής μου:

- Δεν έκανα κάποιο φιλτράρισμα σχετικά με τις αρνητικές ομοιότητες.
- Βρίσκονται οι πρώτες  $k$  όμοιες ταινίες. Αν υπάρχουν λιγότερες, επιλέγονται όσες υπάρχουν.
- Όλες οι όμοιες ταινίες που βρίσκονται έχουν βαθμολογηθεί από τον χρήστη στο train.
- Η 3η συνάρτηση πρόβλεψης (στις εικόνες παρακάτω αναφέρεται ως "Average User Weighted Function") χρησιμοποιεί για τα βάρη την συνάρτηση στάθμισης  $F(x) = 1.06^x$ . Η συγκεκριμένη επιλέχθηκε διότι παρατήρησα ότι συνήθως υπάρχουν πολύ περίπου 120 κοινοί χρήστες, οπότε θεώρησα πως αυτή η συνάρτηση δίνει μικρά βάρη για λίγους χρήστες και αρκετά μεγαλύτερα για πολλούς στο διάστημα που θέλουμε:



Για σταθερό train = 90%, 5 δοκιμές για k (5, 15, 25, 100, 1000):

For k = 5 we have the following scores:

Average Function:

MAE: 0.6537399569989815

Precision: 0.8947076752563037

Recall: 0.8847787368132621

Average Weighted Function:

MAE: 0.6514364532134628

Precision: 0.9011004716306988

Recall: 0.8638169612275655

Average User Weighted Function:

Total predictions:

MAE: 0.6570874626129598

Precision: 0.8975080951710545

Recall: 0.8734073160706947

real 0m52,503s

user 0m48,465s

sys 0m1,510s

For k = 15 we have the following scores:

Average Function:

MAE: 0.643023484416927

Precision: 0.8920815677966102

Recall: 0.9022365072987814

Average Weighted Function:

MAE: 0.6383613738775479

Precision: 0.8952942753720338

Recall: 0.894335074327039

Average User Weighted Function:

Total predictions:

MAE: 0.6408320179366775

Precision: 0.8896750524109015

Recall: 0.9093344047140752

real 0m53,356s

user 0m53,598s

sys 0m1,192s

For k = 25 we have the following scores:

Average Function:

MAE: 0.6490506267177892

Precision: 0.8834347711034666

Recall: 0.9147828469653377

Average Weighted Function:

MAE: 0.6423223182525666

Precision: 0.8878068303094984

Recall: 0.9117687354432114

Average User Weighted Function:

Total predictions:

MAE: 0.6446668516491009

Precision: 0.8811079174287954

Recall: 0.9239621866009042

real 0m57,686s

user 0m52,816s

sys 0m1,465s

For k = 100 we have the following scores:

Average Function:

MAE: 0.6782477446745143

Precision: 0.8731814085232393

Recall: 0.9291683792300315

Average Weighted Function:

MAE: 0.6634008220139546

Precision: 0.8761278680072183

Recall: 0.9312234552678449

Average User Weighted Function:

Total predictions:

MAE: 0.6619741030156063

Precision: 0.8694776589049716

Recall: 0.946431017947664

real 1m12,267s

user 1m9,407s

sys 0m1,598s

For k = 1000 we have the following scores:

Average Function:

MAE: 0.7216848635658272

Precision: 0.8637472710928471

Recall: 0.9214960953555281

Average Weighted Function:

MAE: 0.6889180544142451

Precision: 0.8664631819916062

Recall: 0.9334155363748459

Average User Weighted Function:

Total predictions:

MAE: 0.690304966686529

Precision: 0.863312944701036

Recall: 0.9475270585011646

real 2m20,625s

user 2m18,067s

sys 0m1,572s

Παρατηρούμε ότι τα καλύτερα αποτελέσματα τα έχουμε όταν παίρνουμε τις k=15 πιο όμοιες ταινίες. Δοκιμάζοντας για 3 διαφορετικά ποσοστά του train και με σταθερό 10% test set έχουμε:

➤ Για **train = 50%**:

For k = 15 we have the following scores:

Average Function:

MAE: 0.6909981568110249

Precision: 0.8814569536423841

Recall: 0.8935284640171858

Average Weighted Function:

MAE: 0.6905641209301521

Precision: 0.8850263763019072

Recall: 0.8784908700322234

Average User Weighted Function:

Total predictions:

MAE: 0.6817210514348059

Precision: 0.8814432989690721

Recall: 0.8954081632653061

real 0m51,413s

user 0m51,567s

sys 0m1,268s

➤ Για **train = 70%**:

For k = 15 we have the following scores:

Average Function:

MAE: 0.6633992774873922

Precision: 0.886027469624934

Recall: 0.8984866747020223

Average Weighted Function:

MAE: 0.658362239661748

Precision: 0.8902766586086489

Recall: 0.8877728672827105

Average User Weighted Function:

Total predictions:

MAE: 0.6536632817199454

Precision: 0.885605463619648

Recall: 0.9030400428552297

real 0m51,838s

user 0m52,040s

sys 0m1,225s

➤ Για **train=90%**:

For k = 15 we have the following scores:

Average Function:

MAE: 0.643023484416927

Precision: 0.8920815677966102

Recall: 0.9022365072987814

Average Weighted Function:

MAE: 0.6383613738775479

Precision: 0.8952942753720338

Recall: 0.894335074327039

Average User Weighted Function:

Total predictions:

MAE: 0.6408320179366775

Precision: 0.8896750524109015

Recall: 0.9093344047140752

real 0m53,356s

user 0m53,598s

sys 0m1,192s