

HW4

Import Required Packages

```
library(data.table)
library(tidyr)
library(dplyr)
library(ggplot2)
library(anytime) #install.packages("anytime")
library(recommenderlab) #install.packages("recommenderlab")
library(stringr) #install.packages("stringr")
library(readr)
set.seed(800)
```

Q1.

Read data from file after adding new user

```
#Read the dataset
dataset <- read.csv("cs_books1.csv")
View(dataset)
```

Prepare the data for the recommender system

```
#Prepare the data for the recommender system
ratings_spread <- spread(dataset, key=book, value=rating)
rating_matrix <- as.matrix(ratings_spread[,-1])

dimnames(rating_matrix) <- list(paste("u", unique(dataset$user), sep=""),
                                paste("m", unique(dataset$book), sep=""))
rating_matrix_lab <- as(rating_matrix, "realRatingMatrix")
```

```
#The created Matrix
getRatingMatrix(rating_matrix_lab)
```

```
## 79 x 9 sparse Matrix of class "dgCMatrix"
##               mArtificial intelligence mSystems programming
## uMarvin Minsky                .                1.46
## uDonald Knuth                  2.95                3.75
## uEdgar Codd                    3.44                .
## uMichael Stonebraker           1.11                2.41
## uTony Hoare                    .                2.02
## uJohn Backus                   .                .
## uJohn McCarthy                 1.26                .
## uDennis Ritchie                .                .
## uAlan Perlis                   .                .
## uLeslie Lamport                .                1.68
## uEdsger Dijkstra               1.39                .
## uRobert Floyd                  .                .
## uNiklaus Wirth                 .                3.12
## uRobin Milner                  5.00                .
## uMacauley Mustafa              1.58                3.51
## uSultan Armstrong              .                .
## uLeandro Warner                4.92                .
```

## uJavier Donovan	2.90	.
## uZakariyya Pemberton	3.54	2.68
## uFrancesco Hutton	1.73	4.91
## uJax Sawyer	1.85	4.95
## uLacey-Mai Neal	.	4.28
## uBenny Li	.	.
## uIdris William	1.42	.
## uThierry Truong	4.65	1.09
## uClive Strong	4.41	.
## uMohammod McCabe	.	1.99
## uJasleen Flynn	4.44	.
## uFarrah Traynor	3.45	2.40
## uMilla Callahan	4.63	.
## uHester Dunn	.	3.93
## uSamad Wilkinson	1.88	.
## uNikolas Legge	.	1.13
## uHuma Delaney	.	1.33
## uUrsula Nichols	.	.
## uRoshan Charles	.	.
## uSaima Kearns	4.26	1.19
## uLyndsey Ahmed	.	.
## uFabien Sanchez	3.03	5.00
## uSafiyah Humphrey	.	4.92
## uWilma Sloan	.	4.16
## uLouie Fowler	2.79	3.53
## uZackery Woodcock	4.49	3.20
## uIga Reyes	2.85	.
## uTyler-James Cleveland	4.63	.
## uJoao Hammond	.	2.13
## uEmer Irwin	.	3.62
## uSalma Ellis	2.92	.
## uAntonia Shannon	.	1.69
## uJane Bouvet	.	.
## uEllice Haas	2.76	5.00
## uDrew Peck	2.89	.
## uTymon Sweeney	.	.
## uDarrel Thorne	2.83	2.88
## uKristy Cantrell	3.76	3.19
## uConor Dickson	3.95	.
## uIolo Howell	.	.
## uLloyd Currie	5.00	.
## uShane Mayer	.	.
## uDesiree Riley	.	2.73
## uEmeli Ponce	.	2.20
## uIrfan Heaton	2.00	.
## uKai Plummer	2.03	.
## uXavier Copeland	.	.
## uLeigh Zhang	.	.
## uChyna Kemp	.	2.30
## uLinzi McGill	4.44	1.27
## uFenton Higgs	.	.
## uTed Weber	.	.
## uTeddy Emery	4.38	.
## uAlishba Gaines	4.02	1.17

## uAntoinette Dalby		2.46	2.80
## uMaysa Whyte		2.18	.
## uEshal Philip		.	.
## uAngelika Smart		.	1.68
## uVerity Miles		1.23	.
## uDanyl Connor		.	.
## uHakim Cherry		3.29	.
## uUser1		.	1.57
##	mComputation	mAlgorithms	
## uMarvin Minsky	.	.	
## uDonald Knuth	4.57	2.83	
## uEdgar Codd	.	.	
## uMichael Stonebraker	3.25	.	
## uTony Hoare	.	3.63	
## uJohn Backus	4.70	3.75	
## uJohn McCarthy	.	.	
## uDennis Ritchie	2.36	.	
## uAlan Perlis	.	4.73	
## uLeslie Lamport	.	3.41	
## uEdsger Dijkstra	.	.	
## uRobert Floyd	.	.	
## uNiklaus Wirth	.	4.37	
## uRobin Milner	4.39	.	
## uMacauley Mustafa	2.86	3.49	
## uSultan Armstrong	.	4.28	
## uLeandro Warner	.	3.97	
## uJavier Donovan	4.44	2.28	
## uZakariyya Pemberton	1.04	4.37	
## uFrancesco Hutton	4.06	2.78	
## uJax Sawyer	1.10	4.35	
## uLacey-Mai Neal	2.33	.	
## uBenny Li	1.56	2.04	
## uIdris William	4.22	.	
## uThierry Truong	3.04	.	
## uClive Strong	3.28	3.75	
## uMohammod McCabe	.	.	
## uJasleen Flynn	.	.	
## uFarrah Traynor	4.06	2.78	
## uMilla Callahan	.	.	
## uHester Dunn	2.01	.	
## uSamad Wilkinson	2.05	3.78	
## uNikolas Legge	2.37	2.21	
## uHuma Delaney	.	4.04	
## uUrsula Nichols	4.33	.	
## uRoshan Charles	.	.	
## uSaima Kearns	3.87	2.24	
## uLyndsey Ahmed	.	.	
## uFabien Sanchez	3.23	3.61	
## uSafiyah Humphrey	1.98	4.02	
## uWilma Sloan	.	2.70	
## uLouie Fowler	.	1.09	
## uZackery Woodcock	3.14	.	
## uIga Reyes	4.67	.	
## uTyler-James Cleveland	.	5.00	

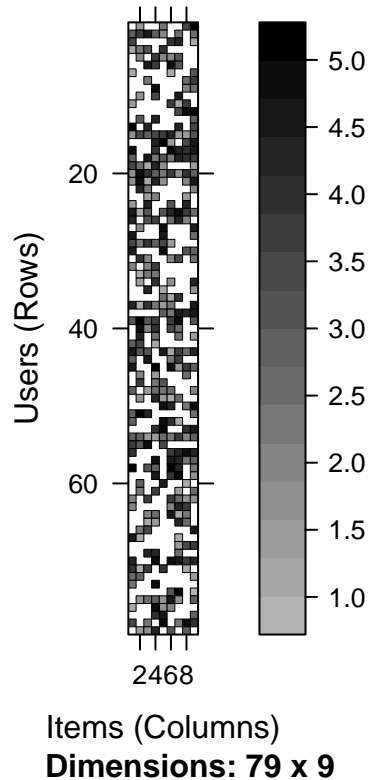
## uJoao Hammond	.	1.48
## uEmer Irwin	.	2.24
## uSalma Ellis	1.48	2.50
## uAntonia Shannon	.	4.37
## uJane Bouvet	.	.
## uEllice Haas	4.32	.
## uDrew Peck	.	.
## uTymon Sweeney	.	4.14
## uDarrel Thorne	1.30	1.54
## uKristy Cantrell	4.99	.
## uConor Dickson	.	.
## uIolo Howell	.	4.77
## uLloyd Currie	3.18	2.92
## uShane Mayer	.	4.62
## uDesiree Riley	.	.
## uEmeli Ponce	.	3.76
## uIrfan Heaton	3.62	.
## uKai Plummer	.	.
## uXavier Copeland	1.86	2.44
## uLeigh Zhang	1.02	1.48
## uChyna Kemp	3.46	.
## uLinzi McGill	.	.
## uFenton Higgs	.	.
## uTed Weber	4.42	4.31
## uTeddy Emery	.	4.88
## uAlishba Gaines	3.61	.
## uAntoinette Dalby	.	.
## uMaysa Whyte	.	4.92
## uEshal Philip	.	.
## uAngelika Smart	2.53	2.78
## uVerity Miles	.	4.85
## uDanyl Connor	.	4.14
## uHakim Cherry	2.26	.
## uUser1	1.74	4.25
##	mProgramming language theory	mConcurrency
## uMarvin Minsky	2.32	.
## uDonald Knuth	1.96	.
## uEdgar Codd	.	4.04
## uMichael Stonebraker	4.03	.
## uTony Hoare	1.86	1.95
## uJohn Backus	.	4.72
## uJohn McCarthy	.	1.12
## uDennis Ritchie	.	.
## uAlan Perlis	.	.
## uLeslie Lamport	.	.
## uEdsger Dijkstra	.	3.82
## uRobert Floyd	.	.
## uNiklaus Wirth	.	.
## uRobin Milner	.	.
## uMacauley Mustafa	3.65	4.40
## uSultan Armstrong	5.00	1.53
## uLeandro Warner	.	5.00
## uJavier Donovan	3.04	2.29
## uZakariyya Pemberton	1.80	.

## uFrancesco Hutton	2.31	3.80
## uJax Sawyer	1.82	.
## uLacey-Mai Neal	.	.
## uBenny Li	.	.
## uIdris William	.	3.28
## uThierry Truong	3.16	4.41
## uClive Strong	.	2.61
## uMohammod McCabe	4.77	.
## uJasleen Flynn	4.47	.
## uFarrah Traynor	3.48	1.48
## uMilla Callahan	4.80	.
## uHester Dunn	.	1.93
## uSamad Wilkinson	.	.
## uNikolas Legge	.	.
## uHuma Delaney	.	1.79
## uUrsula Nichols	1.22	.
## uRoshan Charles	2.48	1.08
## uSaima Kearns	1.30	2.22
## uLyndsey Ahmed	2.80	2.42
## uFabien Sanchez	.	3.58
## uSafiyah Humphrey	.	3.54
## uWilma Sloan	.	.
## uLouie Fowler	.	.
## uZackery Woodcock	4.60	1.02
## uIga Reyes	.	2.40
## uTyler-James Cleveland	.	4.93
## uJoao Hammond	.	.
## uEmer Irwin	3.13	.
## uSalma Ellis	2.19	2.64
## uAntonia Shannon	1.34	1.21
## uJane Bouvet	4.85	4.29
## uEllice Haas	.	.
## uDrew Peck	4.86	3.89
## uTymon Sweeney	5.00	.
## uDarrel Thorne	2.66	2.16
## uKristy Cantrell	3.44	.
## uConor Dickson	.	3.83
## uIolo Howell	.	4.67
## uLloyd Currie	.	5.00
## uShane Mayer	.	3.94
## uDesiree Riley	4.54	.
## uEmeli Ponce	.	.
## uIrfan Heaton	.	3.59
## uKai Plummer	.	1.95
## uXavier Copeland	.	4.51
## uLeigh Zhang	.	.
## uChyna Kemp	.	.
## uLinzi McGill	.	.
## uFenton Higgs	.	.
## uTed Weber	.	.
## uTeddy Emery	.	4.72
## uAlishba Gaines	3.88	.
## uAntoinette Dalby	.	.
## uMaysa Whyte	.	.

## uEshal Philip	.	3.00
## uAngelika Smart	1.63	4.52
## uVerity Miles	1.71	2.10
## uDanyl Connor	4.68	.
## uHakim Cherry	4.76	.
## uUser1	.	3.09
##	mSoftware engineering	mFormal methods
## uMarvin Minsky	.	3.34
## uDonald Knuth	2.59	.
## uEdgar Codd	4.33	.
## uMichael Stonebraker	.	.
## uTony Hoare	3.09	.
## uJohn Backus	3.02	.
## uJohn McCarthy	.	.
## uDennis Ritchie	.	2.43
## uAlan Perlis	.	2.10
## uLeslie Lamport	4.45	.
## uEdsger Dijkstra	1.06	3.70
## uRobert Floyd	3.45	4.83
## uNiklaus Wirth	.	.
## uRobin Milner	4.33	.
## uMacauley Mustafa	1.42	1.45
## uSultan Armstrong	.	3.54
## uLeandro Warner	4.34	4.04
## uJavier Donovan	4.36	3.88
## uZakariyya Pemberton	.	.
## uFrancesco Hutton	1.59	4.06
## uJax Sawyer	.	1.88
## uLacey-Mai Neal	.	.
## uBenny Li	.	1.35
## uIdris William	.	4.24
## uThierry Truong	4.98	3.62
## uClive Strong	4.16	1.94
## uMohammod McCabe	.	.
## uJasleen Flynn	.	.
## uFarrah Traynor	.	.
## uMilla Callahan	.	3.00
## uHester Dunn	3.55	.
## uSamad Wilkinson	.	.
## uNikolas Legge	.	.
## uHuma Delaney	.	.
## uUrsula Nichols	.	.
## uRoshan Charles	3.95	.
## uSaima Kearns	.	.
## uLyndsey Ahmed	4.58	4.38
## uFabien Sanchez	4.72	.
## uSafiyah Humphrey	2.71	.
## uWilma Sloan	2.71	.
## uLouie Fowler	.	.
## uZackery Woodcock	3.83	2.49
## uIga Reyes	3.71	.
## uTyler-James Cleveland	1.50	3.76
## uJoao Hammond	.	2.50
## uEmer Irwin	.	.

## uSalma Ellis	.	.	2.02
## uAntonia Shannon	2.58	.	.
## uJane Bouvet	.	2.17	4.75
## uEllice Haas	.	.	2.54
## uDrew Peck	1.25	.	.
## uTymon Sweeney	.	3.86	4.67
## uDarrel Thorne	3.82	2.05	2.77
## uKristy Cantrell	.	.	.
## uConor Dickson	4.23	4.74	4.22
## uIolo Howell	4.34	2.32	2.86
## uLloyd Currie	4.24	.	2.17
## uShane Mayer	5.00	.	1.66
## uDesiree Riley	.	1.21	.
## uEmeli Ponce	1.52	.	2.59
## uIrfan Heaton	2.49	1.71	1.70
## uKai Plummer	4.23	3.72	.
## uXavier Copeland	1.04	2.39	.
## uLeigh Zhang	4.25	2.60	.
## uChyna Kemp	.	.	4.62
## uLinzi McGill	1.12	.	.
## uFenton Higgs	1.30	3.52	3.64
## uTed Weber	2.53	.	.
## uTeddy Emery	4.64	3.62	4.38
## uAlishba Gaines	.	1.12	.
## uAntoinette Dalby	.	.	1.46
## uMaysa Whyte	.	.	.
## uEshal Philip	5.00	.	1.00
## uAngelika Smart	1.64	.	2.88
## uVerity Miles	.	.	.
## uDanyl Connor	.	.	2.99
## uHakim Cherry	2.20	4.78	3.86
## uUser1	.	2.91	.

```
image(rating_matrix_lab)
```



```
#Print all types of recommenders systems
recommenderRegistry$get_entry_names()
```

```
## [1] "ALS_realRatingMatrix"          "ALS_implicit_realRatingMatrix"
## [3] "ALS_implicit_binaryRatingMatrix" "AR_binaryRatingMatrix"
## [5] "IBCF_binaryRatingMatrix"       "IBCF_realRatingMatrix"
## [7] "POPULAR_binaryRatingMatrix"    "POPULAR_realRatingMatrix"
## [9] "RANDOM_realRatingMatrix"        "RANDOM_binaryRatingMatrix"
## [11] "RERECOMMEND_realRatingMatrix"  "SVD_realRatingMatrix"
## [13] "SVDF_realRatingMatrix"         "UBCF_binaryRatingMatrix"
## [15] "UBCF_realRatingMatrix"
```

```
#get the matrix of the new user
new_user <- rating_matrix_lab[79]
#Select all users except the new user
rating_matrix_lab = rating_matrix_lab[1:78]

rating_matrix_lab
```

```
## 78 x 9 rating matrix of class 'realRatingMatrix' with 357 ratings.
```

```
#Split the dataset into 20% of test data and 80% of train
train_idx <- sample(nrow(rating_matrix_lab), round(nrow(rating_matrix_lab)/100*80,0), replace = F)

train <- rating_matrix_lab[train_idx,]
test <- rating_matrix_lab[-train_idx,]
```


Q2. Predict rating for the first 10 users for each of the 3 recommender systems you have selected in the previous task.

```
#First recommender system
model_popular <- Recommender(train, "POPULAR")
#Predict recommendation for 10 users
prediction_popular <- predict(model_popular, test[1:10], n=10, type="ratings")
as(prediction_popular, "list")
```

```
## $`uTony Hoare`
## mArtificial intelligence      mComputation      mFormal methods
##           2.471337           2.620745           2.254502
##           mDatabases
##           2.374586
##
## $`uDennis Ritchie`
## mArtificial intelligence      mSystems programming
##           2.784671           2.670251
##           mAlgorithms mProgramming language theory
##           3.197917           2.991054
##           mConcurrency      mSoftware engineering
##           2.757587           2.781765
##
## $`uMacauley Mustafa`
## named numeric(0)
##
## $`uLeandro Warner`
## mSystems programming      mComputation
##           4.311917           4.575745
## mProgramming language theory
##           4.632720
##
## $`uMilla Callahan`
## mSystems programming      mComputation      mAlgorithms
##           3.671917           3.935745           4.199584
##           mConcurrency mSoftware engineering
##           3.759254           3.783432
##
## $`uSaima Kearns`
## mSoftware engineering      mFormal methods
##           2.759861           2.545931
##
## $`uSafiyah Humphrey`
## mArtificial intelligence mProgramming language theory
##           3.395337           3.601720
##           mFormal methods      mDatabases
##           3.178502           3.298586
##
## $`uWilma Sloan`
## mArtificial intelligence      mComputation
##           2.718837           2.868245
## mProgramming language theory      mConcurrency
##           2.925220           2.691754
```

```
##           mFormal methods
##           2.502002
##
## $`uIga Reyes`
##           mSystems programming           mAlgorithms
##           3.254417                       3.782084
## mProgramming language theory           mFormal methods
##           3.575220                       3.152002
##           mDatabases
##           3.272086
##
## $`uAntonia Shannon`
## mArtificial intelligence           mComputation           mFormal methods
##           2.199337                       2.348745           1.982502
##           mDatabases
##           2.102586
```

```
#second recommender system
```

```
model_Random <- Recommender(train, "RANDOM")
prediction_Random <- predict(model_Random, test[1:10], n=10, type="ratings")
as(prediction_Random, "list")
```

```
## [[1]]
## mArtificial intelligence           mComputation           mFormal methods
##           1.460000                       3.786047           2.506221
##           mDatabases
##           2.718447
##
## [[2]]
##           mArtificial intelligence           mSystems programming
##           4.166053                       3.055480
##           mAlgorithms mProgramming language theory
##           2.559004                       3.513726
##           mConcurrency           mSoftware engineering
##           3.177326                       3.363046
##
## [[3]]
## named numeric(0)
##
## [[4]]
##           mSystems programming           mComputation
##           3.580131                       4.717481
## mProgramming language theory
##           4.423268
##
## [[5]]
## mSystems programming           mComputation           mAlgorithms
##           3.760331                       3.659947           3.913443
##           mConcurrency mSoftware engineering
##           3.665986                       3.652569
##
## [[6]]
## mSoftware engineering           mFormal methods
##           4.550709                       4.522199
##
```

```

## [[7]]
##      mArtificial intelligence mProgramming language theory
##                      2.438069                      1.040569
##                      mFormal methods                      mDatabases
##                      3.470430                      3.382784
##
## [[8]]
##      mArtificial intelligence                      mComputation
##                      1.092157                      3.586266
## mProgramming language theory                      mConcurrency
##                      1.940291                      2.124737
##                      mFormal methods
##                      3.426962
##
## [[9]]
##      mSystems programming                      mAlgorithms
##                      3.273710                      3.644188
## mProgramming language theory                      mFormal methods
##                      2.486374                      2.694993
##                      mDatabases
##                      4.035116
##
## [[10]]
## mArtificial intelligence                      mComputation                      mFormal methods
##                      1.600178                      2.196812                      3.228939
##                      mDatabases
##                      1.413846

#third recommender system
model_UBCF <- Recommender(train, "UBCF")
prediction_UBCF <- predict(model_UBCF, test[1:10],n=10, type="ratings")
as(prediction_UBCF, "list")

## $`uTony Hoare`
## mArtificial intelligence                      mComputation                      mFormal methods
##                      2.235313                      2.602377                      2.587492
##                      mDatabases
##                      2.512104
##
## $`uDennis Ritchie`
##      mArtificial intelligence                      mSystems programming
##                      2.626102                      2.697487
##                      mAlgorithms mProgramming language theory
##                      3.115956                      3.099870
##                      mConcurrency                      mSoftware engineering
##                      2.632876                      2.743652
##
## $`uMacauley Mustafa`
## named numeric(0)
##
## $`uLeandro Warner`
##      mSystems programming                      mComputation
##                      4.299920                      4.657870
## mProgramming language theory
##                      4.614063

```

```
##
## $`uMilla Callahan`
## mSystems programming      mComputation      mAlgorithms
##          3.738012          3.803740          3.986503
##          mConcurrency mSoftware engineering
##          3.660839          3.865132
##
## $`uSaima Kearns`
## mSoftware engineering      mFormal methods
##          2.852966          2.717202
##
## $`uSafiyah Humphrey`
## mArtificial intelligence mProgramming language theory
##          3.211670          3.663128
##          mFormal methods      mDatabases
##          3.253227          3.552997
##
## $`uWilma Sloan`
## mArtificial intelligence      mComputation
##          2.540003          3.026972
## mProgramming language theory      mConcurrency
##          3.015715          2.733687
##          mFormal methods
##          2.567900
##
## $`uIga Reyes`
## mSystems programming      mAlgorithms
##          3.306265          3.601434
## mProgramming language theory      mFormal methods
##          3.335878          3.392867
##          mDatabases
##          3.477755
##
## $`uAntonia Shannon`
## mArtificial intelligence      mComputation      mFormal methods
##          1.989681          2.321064          2.365532
##          mDatabases
##          2.092464
```

Q3. Use measures (RMSE, MAE) to evaluate the performance of the models. Which model performs better than others and why?

```
eval_scheme <- evaluationScheme(rating_matrix_lab, method="split", train=0.8, given=-1)
#2 ratings of 20% of users (per user) are excluded for testing

model_popular <- Recommender(getData(eval_scheme, "train"), "POPULAR")

prediction_popular <- predict(model_popular, getData(eval_scheme, "known"), type="ratings")

rmse_popular <- calcPredictionAccuracy(prediction_popular,
                                       getData(eval_scheme, "unknown"))[1]
```

```
rmse_popular

##      RMSE
## 1.573969
model_ubcf <- Recommender(getData(eval_scheme, "train"),
                          "UBCF")

prediction_ubcf <- predict(model_ubcf, getData(eval_scheme, "known"), type="ratings")

rmse_ubcf <- calcPredictionAccuracy(prediction_ubcf, getData(eval_scheme, "unknown"))[1]
rmse_ubcf

##      RMSE
## 1.604937
model_random <- Recommender(getData(eval_scheme, "train"), "RANDOM")

prediction_Random <- predict(model_random, getData(eval_scheme, "known"), type="ratings")

rmse_random <- calcPredictionAccuracy(prediction_Random,
                                     getData(eval_scheme, "unknown"))[1]
rmse_random

##      RMSE
## 1.818836
```

Using the following table, we can compare between the different models. We can see that model popular is the best in performance because it has the lowest MSE, RMSE, and MAE. Lowest MSE means that the difference between the original and the predicted one is very low, which means it's a better model.

```
#Compare models

rbind(calcPredictionAccuracy(prediction_popular, getData(eval_scheme, "unknown")),
      calcPredictionAccuracy(prediction_ubcf, getData(eval_scheme, "unknown")),
      calcPredictionAccuracy(prediction_Random, getData(eval_scheme, "unknown")))

##      RMSE      MSE      MAE
## [1,] 1.573969 2.477378 1.336411
## [2,] 1.604937 2.575822 1.363182
## [3,] 1.818836 3.308166 1.600704
```

Q4 Add a new user (with username “User1”) in your data. Suggest what books should User1 read?

We added User1 from the beginning, then we suggest to user1 what he should read.

```
model_popular <- Recommender(getData(eval_scheme, "train"), "POPULAR")
new_userPrediction <- predict(model_popular, new_user, n=10, type="ratings")
as(new_userPrediction, "list")

## $User1
##      mArtificial intelligence mProgramming language theory
##              2.726171              2.818484
```

##	mSoftware engineering	mDatabases
##	2.692743	2.425372