Charles Zheng EE 378b HW 5

==

Setup

```
library(knitr)
opts_knit$set(root.dir=normalizePath('../'))
```

Load the data

```
load('ee378b/ratings.RData')
dim(ratings)
```

```
## [1] 100000      4
```

```
n <- dim(ratings)[1]
```

Form training and test sets

```
train_inds <- sample(n, .8 * n)
train <- ratings[train_inds, ]
test <- ratings[-train_inds, ]
ntr <- dim(train)[1]
nte <- dim(test)[1]
```

Form training matrix

```
n_u <- max(ratings$user)
n_i <- max(ratings$item)
trmat <- matrix(NA, n_u, n_i)
trmat[cbind(train$user, train$item)] <- train$rating
```

1 Use mean of ratings

Using movie mean

```
rmse <- function(y1, y2) sqrt(sum((y1-y2)^2)/length(y1))
means_i <- colMeans(trmat, na.rm = TRUE)
pr_ratings <- means_i[test$item]
pr_ratings[is.na(pr_ratings)] <- mean(train$rating)
rmse_mean_movie <- rmse(pr_ratings, test$rating)
rmse_mean_movie
```

```
## [1] 1.031033
```

Using user mean

```
means_u <- rowMeans(trmat, na.rm = TRUE)
pr_ratings <- means_u[test$user]
pr_ratings[is.na(pr_ratings)] <- mean(train$rating)
rmse_mean_user <- rmse(pr_ratings, test$rating)
rmse_mean_user
```

```
## [1] 1.041624
```

2 Use SVD

Center training matrix by means

```
trmat_c <- t(t(trmat) - means_i)
trmat_c[is.na(trmat_c)] <- 0
library(rARPACK)
res <- svds(trmat_c, k = 10)
pr_k10 <- res$u %*% diag(res$d) %*% t(res$v)
dim(pr_k10)
```

```
## [1]  943 1682
```

```
adj_k10 <- pr_k10[cbind(test$user, test$item)]
```