5.17 The recosystem Package

The **recosystem** package does matrix factorization specifically for recommender systems, i.e. specifically for settings in which the matrix A has many missing values. It's written by experts in numerical matrix factorization, and features a number of useful options.

Below is a **recosystem** session using the small MovieLens data. Let's suppose we've already decided on rank $k = 20$, say by cross validation, and now we'll go back to using the full dataset for our predictions.

```
> library(recosystem)
# recosystem uses 'R6' R objects; all action will take place within r;
# typically the output of a function will be stored as a new component in r
> r <- Reco()
> ml <- read.table('u.data',header=F)
# need to create an object of class 'DataSource', specifying which
# columns are user IDs, item IDs and ratings
> ml.trn <- data_memory(ml[,1],ml[,2],ml[,3],index1=TRUE)

# do the factorization, with rank 20; do use NMF
> r$train(ml.trn,opts=list(dim=20,nmf=TRUE))
iter      tr_rmse          obj
   0       2.0381   5.0056e+05
   1       1.0296   1.7402e+05
   2       0.9529   1.6028e+05
   3       0.9449   1.5868e+05
   4       0.9418   1.5811e+05
   5       0.9397   1.5774e+05
   6       0.9382   1.5749e+05
   7       0.9371   1.5729e+05
   8       0.9362   1.5713e+05
   9       0.9355   1.5701e+05
  10       0.9348   1.5690e+05
  11       0.9343   1.5681e+05
  12       0.9338   1.5673e+05
  13       0.9334   1.5666e+05
  14       0.9330   1.5660e+05
  15       0.9327   1.5654e+05
  16       0.9324   1.5649e+05
  17       0.9321   1.5645e+05
  18       0.9318   1.5641e+05
  19       0.9316   1.5637e+05
```

```
# training went for 20 iterations; RMSE is the square root
#    of mean squared error
# for large data, write to disk, otherwise in memory
> result <- r$output(out_memory(),out_memory())
> str(result)
List of 2
 $ P: num [1:943, 1:20] 0.676 0.677 0.574 0.836 0.574 ...
 $ Q: num [1:1682, 1:20] 0.712 0.614 0.568 0.645 0.612 ...
# P and Q are W and H'
> w <- result$P
> h <- t(result$Q)
# let's try a prediction, with a known rating
> head(ml)
   V1  V2 V3        V4
1 196 242  3 881250949
2 186 302  3 891717742
3  22 377  1 878887116
...
> w[22,] %*% h[,377]
         [,1]
[1,] 2.196976
# or just have recosystem do it for us
> preds <- r$predict(ml.trn,out_memory())
> head(preds)
[1] 3.979107 4.212397 2.196976 3.601082 3.900878 4.467487
```