Null: Final Presentation

Xintong Lin, Darius Simmons, Nick Cho

Models and (best) RMSE's

- SVD
  - 0.89342 (1000 features, 23 epochs)
- SVD++
  - 0.89167 (200 features, 50 epochs)
- Time SVD++
  - 0.88216 (800 features, 80 epochs)
- RBM
  - 0.92281 (75 features, 50 epochs)
- KNN
  - 0.94931 (Pearson, Residuals, Number of Viewers)
- Weekday
  - 1.56002

Blending

- Final blend using ridge regression included:
  - SVD (13)
  - SVD++ (15)
  - Time SVD++ (17)
  - RBM (9)
  - KNN (5)
  - Other blends (3) - including 2 neural network blends
  - Weekday (2)

What worked well

- Armadillo (c++ linear algebra library)
  - used for SVD and allowed for each epoch to be completed in ~1min
- CMake
- Preprocessing movie correlations for KNN
- Training at the end on base+probe data gave ~0.08 improvement in quiz RMSE

What did not work well

- Armadillo (c++ linear algebra library)
  - not used for SVD++/SVD++ time so each epoch with features >100 completed in >10mins
- Trying to over optimize KNN, ended up working on the smaller dataset but not on the full data
- Did not see an improvement in conditional RBM over regular RBM, and conditional RBM took longer to run

What we learned

- Space and Time complexity become very important
- Balance between performance and diversity for blending

Results

- Final RMSE on quiz: 0.87193 (8.353% above water)
- RMSE on test: 0.87282 (8.3654% above water)

Final Thoughts

- Used the models and techniques described by BigChaos, BellKor, etc.
- With more time, would have tried more novel algorithms