Baseline Embedding - slide 5 and 6 (1 minute)

* Hi im annie and I’ll give a brief overview of our baseline model for the movie recommendation task
* Our baseline model is a vanilla embedding model with the number of embedding dimensions of 100 that conducts collaborative filtering through a matrix factorization computation between a user and an item matrix.
* With this baseline mode, we achieved a minimum RMSE at epoch 30 of 0.8773, so the leading validation loss was 0.7627.
* We were also able to use this model to generate ratings prediction and movie recommendations for user with the ID 2000
* Next, Mike will talk about some tuning we did for our baseline model

Autoencoders - slide 16, 17, 18 (2 mins)

* Thanks drew, Hi its annie again and i’ll be going over the autoencoder application of our project
* Autoencoders are a type of neural networks that learns a representation by dimensionality reduction and noise elimination
* AutoRec is a compact and efficiently trainable autoencoder for collaborative filtering (CF) in recommendation tasks. It was created by a group of researchers at the Australian National University in 2015, which at the time, had outperformed state-of-the-art CF techniques
* In our project, we focused on conducting a deep learning re-implementation of AutoRec based off of Zheda Mai’s approach, who’s a graduate researcher in Recommender Systems at the University of Toronto
* First, we preprocess our data to meet the structure required by the autoencoder
* we start of with split the data into train est val ratified the data by user\_id to ensure class balances
* transform the data in these different sets into the user-item matrix with the number of rows = number of users, number of columns = number of movies
* set many default values for unobserved ratings, which can take values from 0 to 5 → to investigate its effect on RMSE scores
* Also had a customed masked RMSE function bc we only want to take into accounts ratings that arent 0
* Next slide pls, Built 2 models, tuned many key hyperparameters, and tried multiple other approaches, such as training on the different training matrices with different rating values, loading in the larger 10M movielens dataset (though that didn’t really work out), adding noise to the model, adding more predictive demographic features to the model and implementing a demographic-based recommender system approach in which we train the model on a subset of the most popular demographic groups: male of age 18 - 24 years old and predict ratings/ recommend movies for a user in this demographic subset
* Next slide pls, After multiple approaches, we have come to conclusion that this is our best model the Deep AutoRec model trained on the full 6040 x 3952 training matrix with default rating set to 0 with the set of parameters(: layers = [256, 512, 256], dropout = 0.8, activation and last\_activation = 'selu’, and regularization alpha of encoder and decoder = 0.001.) We achieved a masked RMSE of 0.858 and loss of 0.9957, achieving a slight improvement compared to our vanilabe embedding baseline
* Next mike will be going over word2vec content filtering approach