

Restricted Boltzmann Machines

Introduction on problem and technique

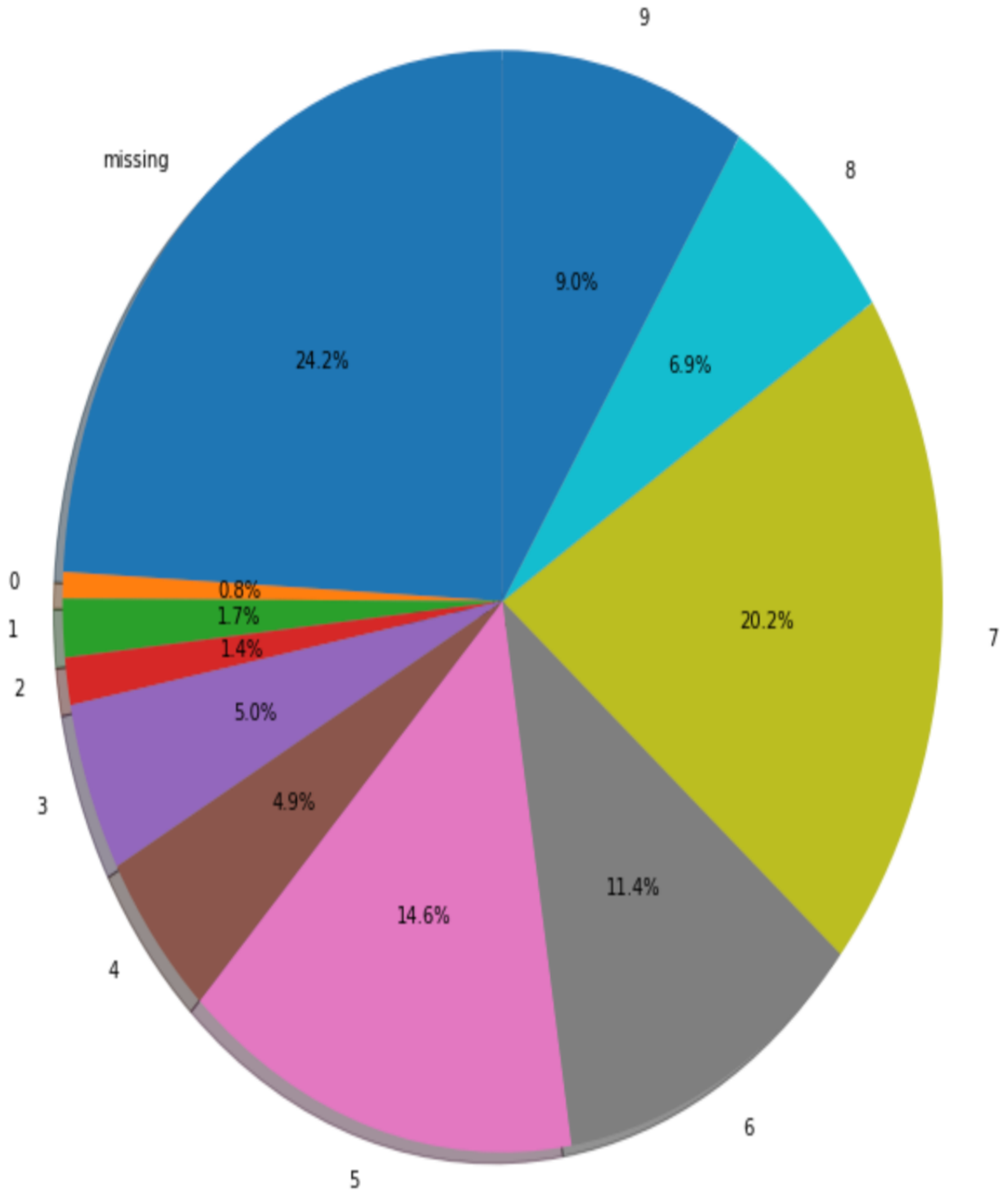
Recommendation systems is an important information filter used in a variety of areas. It is commonly used for playlist generators for video and music services, product recommenders for online stores, or content recommenders for social media platforms and open web content recommenders. Therefore, we are interested in the system and trying to create our own version.

We use packages like BernoulliRBM, pipeline from sklearn to generate a Restricted Boltzmann Machines model and modules such as numpy, pandas and matplotlib to visualize and analyze our dataset and result quantitatively. In addition, we not only used the processed datasets provided by instructor from Canvas, but also adapted some of the original datasets from Kaggle: MovieLens 20M Dataset

Read Data and visualization

We read data from dataset, and did a little visualization to know a basic distribution of dataset.

```
with open('mlens-small/top-names.txt') as f: names =  
f.read().split('\n')  
  
ratings = np.loadtxt('mlens-small/top-ratings-missing.txt')  
  
nUsers, nMovies = ratings.shape  
  
X = (ratings >= 7).astype(int); # set movies ratings over 7 as  
recommandated
```



This Pie Chart shows that from the raw data, we have a lot of missing ratings, and ratings over 5 share a big portion.