Lab Course Machine Learning
Exercise 8

Prof. Dr. Dr. Lars Schmidt-Thieme,
Mofassir ul Islam Arif
Information Systems and Machine Learning Lab
University of Hildesheim
Submission: 20.01.2019 LearnWeb 3112

January 19, 2020

Instructions

Please following these instructions for solving and submitting the exercise sheet.

1. You should submit two things a) python scripts(zipped) / jupyter notebook and b) a pdf document.
2. In the pdf document you will explain your approach (i.e. how you solved a given problem), and present your results in form of graphs and tables.
3. The submission should be made before the deadline, only through learnweb.
4. Unless explicitly mentioned, you are not allowed to use scikit, sklearn or any other library for solve any part. All implementations must be done yourself.

1 Exercise Sheet 10

Datasets

- 1. Recommender Datasets: You can use one of the two datasets (or optionally, both datasets).

  (a) movielens 100k dataset D1: Rating prediction dataset (rating scale 1-5). http://grouplens.org/datasets/movielens/100k/

- 2. The RMSE score for rating prediction is available at Mymedialite website http://www.mymedialite.net/examples/datasets.html

Exercise 1: Recommender Dataset (4 Points)

Perform the statistical analysis of the two datasets given. Your analysis should provide as much information as possible. You must use all the related information of users and movies for the analysis i.e. rating, user (age group, zipcode etc) and item(genre, title, release date etc). The grading of this task depends on the useful information extracted from the given datasets, which can help in the learning process. Use tables, graphs to represent your information.

Exercise 2: Implement basic matrix factorization (MF) technique for recommender systems (8 Points)

In this task you are required to implement a matrix factorization (MF) technique for recommender systems (see Annex 1). You are given a rating matrix $R^{n \times m}$ and you have to learn latent matrices $P^{n \times k}$ and $Q^{m \times k}$, where n is the number of users, m is the number of items and k the latent dimensions. You can solve the MF problem by implementing Stochastic Gradient Descent (SGD) or Alternating Least Square(ALS) or Coordinate Descent(CD) learning algorithms. Measure the prediction quality (the RMSE score) on the test dataset.

- normalize your data
- optimize the hyper-parameters i.e. λ regularization constant, α learning rate, k latent dimensions.
- Compute the test RMSE.

Exercise 3: Recommender Systems using matrix factorization libmf / sckit-learn (8 Points)

In this task you are required to use off-the-shelf libraries such as libmf or sckit-learn. You have to learn a matrix factorization model using coordinate descent method. Optimize the hyper parameters and perform a 3-fold cross validation. Compare your results with the results in task 1. List in detail which/how you used these libraries?, what it solves?, and why it is selected?. Present your results in form of plots and tables.

1.1 ANNEX

- Latent Feature Recommender System `https://www.youtube.com/watch?v=E8aMcwmqsTg`
- Matrix Factorization Technique for Recommender System: by Y. Koren, https://datajobs.com/data-science-repo/Recommender-Systems-%5BNetflix%5D.pdf
- sklearn.model_selection, sklearn.metrics, sklearn.linear_model, sklearn.preprocessing
- Scikit Learn User Guide http://scikit-learn.org/stable/user_guide.html
- You can use matplotlib for plotting.
- sklearn.metrics http://scikit-learn.org/stable/modules/classes.html#module-sklearn.metrics