Recommender Systems

Reading

- In case you are further interested in Recommender Systems than what we cover in this course, I suggest you take a look at:
    - Recommender Systems by Jannach and Zanker
- It is a great textbook on this topic if you want to get a deeper understanding of recommender systems.

Recommender Systems

- Fully developed and deployed recommendation systems are extremely complex and resource intensive
- We will begin to expand beyond our local computer and show you how to use Databrick’s Notebook Platform
- Databricks is a company started by the creators of Spark
- Databricks is constantly maintaining and adding new features to Spark, as well as hosting a data platform.

Recommender Systems

- The Databricks platform provides a very convenient way to quickly create a large Spark cluster that you can interact with in a Notebook Environment
- It’s also ideally set up to work with DataFrames as it has the capability to upload table-like data from a variety of sources

Recommender Systems

- In this section of the course we will combine learning how to create Recommender Systems with learning how to use the Databricks Platform.
- Databricks is also one of the leading providers of official Spark certification, so it may be an added benefit to understand their platform
- Luckily they also have a 6GB free tier!

Recommender Systems

- Since full recommender systems require a heavy linear algebra background we will try to provide only a high level overview in this lecture.
- Check out the book mentioned in the beginning of the lecture for a deeper look into this topic

Recommender Systems

The two most common types of recommender systems are **Content-Based** and **Collaborative Filtering (CF)**.

- Collaborative filtering produces recommendations based on the knowledge of users' attitude to items, that is it uses the "wisdom of the crowd" to recommend items.
- Content-based recommender systems focus on the attributes of the items and give you recommendations based on the similarity between them.

Recommender Systems

- In general, Collaborative filtering (CF) is more commonly used than content-based systems because it usually gives better results and is relatively easy to understand (from an overall implementation perspective).
- The algorithm has the ability to do feature learning on its own, which means that it can start to learn for itself what features to use.

Recommender Systems

- These techniques aim to fill in the missing entries of a user-item association matrix.
- spark.ml currently supports model-based collaborative filtering, in which users and products are described by a small set of latent factors that can be used to predict missing entries.
- spark.ml uses the alternating least squares (ALS) algorithm to learn these latent factors.

Recommender Systems

- ALS is basically a Matrix Factorization approach to implement a recommendation algorithm you decompose your large user/item matrix into lower dimensional user factors and item factors.
- To fully understand this model you need to have a very strong background in Linear Algebra
- Check out the various resource links for more detail on ALS and how it works

Recommender Systems

- For this course, we will leave further review of ALS mathematics up to the student and those resource links.
- We will move on to setting up the Databricks Platform and running a recommendation system on it.
- Let’s get started!