PART I -
UNDERSTANDING RECOMMENDER SYSTEMS

Answer the following questions:

1. Recommender Systems (RS) are based on one of two strategies – content filtering and collaborative filtering. Write down key features and properties of each of them.

Content Filtering

- Creates a profile for each user or product to characterize its nature.
- Require gathering external information which might not be available or easy to collect.

Collaborative Filtering

- Relies on only past user behavior, no need to create profiles.
- Analyzes relationships between users and interdependencies among products to identify new user-item associations.
- It is domain free, yet it can address data aspects that are often elusive and difficult to profile using content filtering.
- More accurate than content-based filtering.
- Suffers from 'cold-start problem', due to its inability to address the system's new products and users.

2. Two primary areas of collaborative filtering are neighborhood methods and latent factor models. Write down key features of each of them.

Neighborhood Methods

- Centered on computing the relationships between items or alternatively between users.
- The item oriented approach evaluates a user's preference for an item based on ratings of "neighboring" items by the same user. A product's neighbors are other products that tend to get similar ratings when rated by the same user.

Latent Factor Models

- Tries to explain the ratings by characterizing both items and users on, say, 20 to 100 factors inferred from the ratings patterns.

3. What do the "factors" in latent factor represent? How are these factors discovered?

- Factors are features based on which the Users and Items are characterized. For example, for item 'movies' the discovered factors might measure obvious dimensions such as comedy vs drama, amount of action, or orientation to children. For users, each factor measures how much the user likes movies that score high on the corresponding movie factor.

- Factors in the latent factor model are discovered from the ratings patterns.

4. What is the difference between explicit and implicit feedback for RS? What do you think are advantages and disadvantages of each?

Difference between Explicit Feedback and Implicit Feedback

- Explicit feedback is represented by a sparse matrix since any single user is likely to have rated only a small percentage of possible items, whereas Implicit feedback is represented by a densely filled matrix.

Advantages of Explicit Feedback

- High accuracy as it is the ratings directly from the users.

Disadvantages of Explicit Feedback

- Sparsely filled matrix.

Advantages of Implicit Feedback

- Densely filled matrix.

Disadvantages of Implicit Feedback

- Low in accuracy as suggested by the Recommender system itself.

5. In the latent factor model for dimension f, the item i and user u are represented as vectors q_i and $p_u \in R_f$. How is the interaction between item i and user u calculated?

- The resulting ***dot product, $q_i^T p_u$***, captures the interaction between user u and item i—the user's overall interest in the item's characteristics.

6. What does the learning system try to minimize? Understand the meaning of each term in the equation.

- To learn the factor vectors (pu and qi), the system minimizes the ***regularized squared error*** on the set of known ratings.

$$\min_{q^*,p^*} \sum_{(u,i)\in K} (r_{ui} - q_i^T p_u)^2 + \lambda(|| q_i ||^2 + || p_u ||^2) \qquad (2)$$

r_{ui} = Approximation of user u's rating for item i
q_i = Each item i is associated with a vector $q_i \in R_f$
p_u = Each user u is associated with a vector $p_u \in R_f$
κ = is the set of the (u,i) pairs for which rui is known (the training set).
λ = controls the extent of regularization and is usually determined by cross-validation

7. There are 2 learning algorithms for latent factorization – stochastic gradient descent (SGD) and alternating least squares (ALS). What are the advantages of ALS over SGD?

Advantages of ALS over SGD –

- ALS can be used in a parallelized system as it computes each q_i independently of the other item factors and computes each p_u independently of the other user factors.
- ALS can also be used for systems centered on implicit data. Because the training set cannot be considered sparse, looping over each single training case—as gradient descent does—would not be practical. ALS can efficiently handle such cases.

8. Read about the Netflix competition and the authors' entry. What were the most descriptive dimensions (features) that their models discovered? Summarize briefly. Also mention what metric do they use to check the performance of their models.

- The first factor vector discovered from the Netflix user-movie matrix has one side lowbrow comedies and horror movies aimed at male or adolescent audience while the other side contains drama or comedy with serious undertones and strong female leads. The second factorization has independent, critically acclaimed, quirky films on the top and on the bottom, mainstream formulaic films.
- They used RMSE as the metric to measure the performance of their models.