Singular Value Decomposition for Recommendation

(1) Let $\underline{\underline{R}} = [R_{ij}]$ an $m \times n$ rating matrix where

$R_{ij} \in \{0, 1, 2, 3, 4, 5\}$ where the 0 value represents the absence of rating. Moreover, m denotes the number of users and n the number of items within a given database.

(2) The general aim of SVD is to determine the matrices $\underline{\underline{U}} \in \mathcal{M}_{k \times m}$ and $\underline{\underline{V}} \in \mathcal{M}_{k \times n}$ such that the original rating matrix $\underline{\underline{R}}$ can be written in the following form:

$$\boxed{\underline{\underline{R}} \approx \underline{\underline{U}}^T \underline{\underline{V}} \quad [1]}$$

where $\underline{\underline{U}}$ is the users' matrix and $\underline{\underline{V}}$ is the items' matrix.

(3) The decomposition problem outlined be Eq. 1 can be formally defined as:

$$\boxed{\min_{\underline{\underline{U}}, \underline{\underline{V}}} \frac{1}{2} \| \underline{\underline{R}} - \underline{\underline{U}}^T \underline{\underline{V}} \|_F^2 \quad [2]} \text{ where } \| x \|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}^2$$

(4) The optimizatio problem defined in Eq. 2 cannot be solved when $\underline{\underline{R}}$ contains un-rated items. This entails, that the optimization problem must be re-formulated as:

$$\boxed{\min_{\underline{\underline{U}}, \underline{\underline{V}}} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} (R_{ij} - \langle U_i \cdot V_j \rangle)^2 + \frac{\lambda_1}{2} \| \underline{\underline{U}} \|_F^2 + \frac{\lambda_2}{2} \| \underline{\underline{V}} \|_F^2 \quad [3]}$$

Eq. 3 can be re-written as:

$$\min_{\underline{\underline{U}},\underline{\underline{V}}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij}-U_i^T V_j)^2 + \frac{\lambda_1}{2}\sum_{i=1}^{m}\|U_i\|^2 + \frac{\lambda_2}{2}\sum_{j=1}^{n}\|V_j\|^2 \quad [4]$$

where $\underline{\underline{I}}$ serves as an indicator function such that:

$$I_{ij} = \begin{cases} 1, & R_{ij} \neq 0\,; \\ 0, & \text{otherwise} \end{cases} \quad [5]$$

Finally, Eq. 4 is re-written as:

$$\min_{\underline{\underline{U}},\underline{\underline{V}}} \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij}-U_i^T V_j)^2 + \frac{\lambda_1}{2}\sum_{i=1}^{m}U_i^T U_i + \frac{\lambda_2}{2}\sum_{j=1}^{n}V_j^T V_j \quad [6]$$

Therefore, by defining the objective function to be minimized $f(\underline{\underline{U}},\underline{\underline{V}})$ as

$$f(\underline{\underline{U}},\underline{\underline{V}}) = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}(R_{ij}-U_i^T V_j)^2 + \frac{\lambda_1}{2}\sum_{i=1}^{m}U_i^T U_i + \frac{\lambda_2}{2}\sum_{j=1}^{n}V_j^T V_j \quad [7]$$

the optimization problem becomes:

$$\min_{\underline{\underline{U}},\underline{\underline{V}}} f(\underline{\underline{U}},\underline{\underline{V}}) \quad [8]$$

⑤ Solving the optimization problem defined in Eq. 8 by utilizing gradient descent we need to evaluate the following expressions:

$$\frac{\partial}{\partial \underline{u}_i} f(\underline{\underline{U}}, \underline{\underline{V}}), \forall i \in [m] \text{ and } \frac{\partial}{\partial \underline{v}_j} f(\underline{\underline{U}}, \underline{\underline{V}}), \forall j \in [n]$$

which are equivalent to:

$$\boxed{\frac{\partial}{\partial \underline{u}_i} f(\underline{\underline{U}}, \underline{\underline{V}}) = \nabla_{\underline{u}_i} f(\underline{\underline{U}}, \underline{\underline{V}}), \quad \forall i \in [m] \quad [9]}$$

$$\boxed{\frac{\partial}{\partial v_j} f(\underline{\underline{U}}, \underline{\underline{V}}) = \nabla_{\underline{v}_j} f(\underline{\underline{U}}, \underline{\underline{V}}), \quad \forall j \in [n] \quad [10]}$$

The final evaluation of expressions [9] and [10] yields that:

$$\nabla_{\underline{u}_i} f(\underline{\underline{U}}, \underline{\underline{V}}) = \lambda_1 \underline{u}_i + \sum_{j=1}^{n} I_{ij} (\underline{u}_i^T \underline{v}_j - R_{ij}) v_j \quad [11] \quad \forall i \in [m]$$

$$\nabla_{v_j} f(\underline{\underline{U}}, \underline{\underline{V}}) = \lambda_2 v_j + \sum_{i=1}^{m} I_{ij} (u_i^T v_j - R_{ij}) u_i \quad [12] \quad \forall j \in [n]$$

⑥ The gradient descent-based algorithm for determining the optimal values $\underline{u}_i^*$, $\forall i \in [m]$ and $\underline{v}_j^*$, $\forall j \in [n]$ would be the following:

- Repeat until convergence:
  - $\underline{u}_i(t+1) = \underline{u}_i(t) - \gamma \nabla_{\underline{u}_i(t)} f(\underline{\underline{U}}(t), \underline{\underline{V}}(t))$
  - $\underline{v}_j(t+1) = \underline{v}_j(t) - \gamma \nabla_{\underline{v}_j(t)} f(\underline{\underline{U}}(t), \underline{\underline{V}}(t))$