Lecture 14-15

- Collaborative Filtering
  (Model-Based Approaches)

IT492: Recommendation Systems (AY 2023/24) — **Dr. Arpit Rana**

Flaws in Memory-Based Collaborative Methods

Memory-based methods rely on rating correlation and have the following flaws.

- These methods assume that users can be neighbors only if they have rated common items.
    - This assumption is very limiting, as users having rated a few or no common items may still have similar preferences.

Flaws in Memory-Based Collaborative Methods

Memory-based methods rely on rating correlation and have the following flaws.

- Since only items rated by neighbors can be recommended, the (catalog) coverage of such methods can also be limited.

Flaws in Memory-Based Collaborative Methods

Memory-based methods rely on rating correlation and have the following flaws.

- These methods suffer from (or are sensitive to) the lack of available ratings (a.k.a. ***sparsity***).
    - Users or items newly added to the system may have no ratings at all, a problem known as ***cold-start***.

Learning-Based Collaborative Methods

Learning-based methods obtain the similarity or affinity between users and items

- by defining a parametric model that describes the relation between users, items or both, and then
- computes the model parameters through an optimization process.

Learning-Based Collaborative Methods

Learning-based methods have a few advantages over memory-based methods.

- These methods can capture high-level patterns and trends in the data, are generally more robust to outliers,
- They are known to generalize better than approaches solely based on local relations.
- These methods require less memory because the relations between users and items are encoded in a limited set of parameters.
- Since the parameters are usually learned offline, the online recommendation process is generally faster.

Learning-Based Collaborative Methods

Learning-based methods that use neighborhood or similarity information can be divided in two categories:

- **Factorization methods (e.g. MF),** and
- Adaptive neighborhood learning methods (e.g. SLIM).

Learning-Based Collaborative Methods

Factorization methods

- These methods project users and items into a reduced latent space that captures their most salient features.
- A relation between two users can be found, even though these users have rated different items, thus, are generally less sensitive to sparse data.
- There are essentially two ways in which factorization can be used:
  - Factorization of a sparse ***similarity matrix***, and
  - Factorization of a sparse **user-item rating matrix**.

- Concepts of Linear Algebra

Eigenvalues and Eigenvectors

$$\boxed{A v = \lambda v}$$

where $A \in \mathbb{R}^{m \times m}$ (Square Matrix)

eigenvectors $\longrightarrow$ $v \in \mathbb{R}^{m \times 1}$ (Column Vector)

eigenvalues $\longrightarrow$ $\lambda \in \mathbb{R}^{m \times m}$ (Diagonal Matrix)

Eigenvalues and Eigenvectors

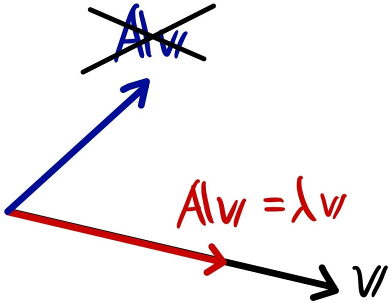
$$\boxed{A v = \lambda v}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}$$

Eigenvalues and Eigenvectors

Eigenvectors are the vectors that does not change its orientation, but just scales by a factor of its corresponding eigenvalue.

$$\boxed{A v = \lambda v}$$



Matrix Transformation does not change the original vector v. Just scales the vector!

Eigenvalues and Eigenvectors

Eigenvectors are the vectors that does not change its orientation, but just scales by a factor of its corresponding eigenvalue.

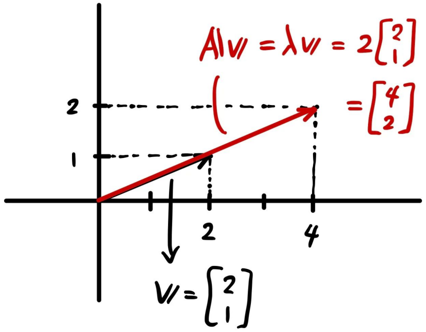
No matter what kind of transition matrix “A” you have, if you managed to find its eigenvalues and eigenvectors, the transition using the matrix “A” on eigenvectors does not change its direction, but just scales by a factor of the corresponding eigenvalues.

$$Av = \lambda v$$

(i) $\lambda_1 = 2$

$$\begin{bmatrix} 1 & 2 \\ 3 & -4 \end{bmatrix}\begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2+2 \\ 6-4 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

$$= \lambda_1 v$$



Diagonalization

- Must be a Square matrix
- Has to have linearly independent eigenvectors

$$\boxed{S^{-1} A S = \Lambda}$$

where
- $S \in \mathbb{R}^{m \times m}$
  - ↳ Block matrix
    Columns are eigenvectors of A
- $A \in \mathbb{R}^{m \times m}$
  - ↳ Square Matrix of interest
- $\Lambda \in \mathbb{R}^{m \times m}$
  - ↳ Diagonal Matrix
    diagonal elements are the eigenvalues of A

Diagonalization

Derivation of Diagonalization

Suppose we have m linearly independent eigenvectors of $\mathbb{A}$.

$$\mathbb{A}\mathbb{S} = \mathbb{A}\begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} = \begin{bmatrix} \mathbb{A}v_1 & \mathbb{A}v_2 & \cdots & \mathbb{A}v_m \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_m v_m \end{bmatrix}$$

(v_i : eigenvector in a column vector)

($\because \mathbb{A}v = \lambda v$)

$$= \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \underbrace{\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{bmatrix}}$$

Diagonal Matrix with eigenvalues in the diagonal elements.

$$\mathbb{A}\mathbb{S} = \mathbb{S}\Lambda$$

Diagonalization and Eigendecomposition

From

$$AS = S\Lambda \xrightarrow[\text{from the left}]{\text{multiply by } S^{-1}} S^{-1}AS = S^{-1}S\Lambda = \Lambda$$

$$\Big\downarrow \text{multiply by } S^{-1} \text{ from the right}$$

$$ASS^{-1} = S\Lambda S^{-1} \longrightarrow \boxed{A = S\Lambda S^{-1}}$$

Diagonalization and Eigendecomposition

$$\boxed{S^{-1} A S = \Lambda}$$

↳ Assumption : S is invertible.

Must have linearly independent eigenvectors.

$A = S \Lambda S^{-1}$ is very useful !

Example

$$A^2 = (S \Lambda S^{-1})(S \Lambda S^{-1})$$

$$= S \Lambda \underbrace{S^{-1} S}_{\text{Identity matrix}} \Lambda S^{-1} = S \Lambda^2 S^{-1}$$

In General, $A^k = S \Lambda^k S^{-1}$

Symmetric Matrix

It is a matrix that comes back to its own when transposed.

$$\boxed{A^T = A}$$

Example

$$A = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 2 & 0 \\ -1 & 0 & 4 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & -1 \\ 3 & 2 & 0 \\ -1 & 0 & 4 \end{bmatrix}$$

Symmetric

Flip based on the diagonal line.
If it's the same as before, it's Symmetric.

$$B = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 3 & -1 \\ 2 & -1 & 5 \end{bmatrix} \qquad B^T = \begin{bmatrix} 1 & \boxed{4} & \boxed{2} \\ \boxed{2} & 3 & -1 \\ \boxed{4} & -1 & 5 \end{bmatrix}$$

Not Symmetric

Check these □ sections.

$$B \neq B^T$$

Symmetric Matrix

It is a matrix that comes back to its own when transposed.

$$\boxed{A^T = A}$$

If a matrix is symmetric,

1. **the eigenvalues are REAL (not complex numbers),** and
2. the eigenvectors could be made perpendicular (orthogonal to each other).

Proof for 1

$$A\mathbf{x} = \lambda \mathbf{x} \xrightarrow{\text{conjugate}} A\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}$$

$$\overline{a+jb} = a - jb$$

Multiply $\bar{\mathbf{x}}^T$ from left

Transpose

$$\bar{\mathbf{x}}^T A^T = \bar{\mathbf{x}}^T \bar{\lambda}$$

$$\bar{\mathbf{x}}^T A = \bar{\mathbf{x}}^T \bar{\lambda} \quad (\because A^T = A)$$

$$\boxed{\bar{\mathbf{x}}^T A \mathbf{x} = \bar{\mathbf{x}}^T \lambda \mathbf{x}} \qquad \boxed{\bar{\mathbf{x}}^T A \mathbf{x} = \bar{\mathbf{x}}^T \bar{\lambda} \mathbf{x}}$$

(Multiplied by $\mathbf{x}$ from right)

Compare

$$\lambda = \bar{\lambda} \qquad \therefore \lambda \text{ is REAL.}$$

Symmetric Matrix

It is a matrix that comes back to its own when transposed.

$$\boxed{A^T = A}$$

If a matrix is symmetric,

1. the eigenvalues are REAL (not complex numbers), and
2. **the eigenvectors could be made perpendicular (orthogonal to each other).**

Proof for ②

Inner product of v, w

$$\langle v, w \rangle = v^T w$$

If A is Symmetric,

$$\langle Av, w \rangle = \langle v, Aw \rangle$$

because

$$\begin{aligned}\langle Av, w \rangle &= (Av)^T w \\ &= v^T A^T w \\ &= v^T A w \quad (\because A^T = A) \\ &= \langle v, Aw \rangle\end{aligned}$$

Symmetric Matrix

It is a matrix that comes back to its own when transposed.

$$\boxed{A^T = A}$$

If a matrix is symmetric,

1. the eigenvalues are REAL (not complex numbers), and
2. **the eigenvectors could be made perpendicular (orthogonal to each other).**

$\lambda_i \neq \lambda_j$

$$\begin{aligned}
\lambda_i \langle x_i, x_j \rangle &= \langle \lambda_i x_i, x_j \rangle \\
&= \langle A x_i, x_j \rangle \quad (\because A x_i = \lambda_i x_i) \\
&= \langle x_i, A x_j \rangle \\
&= \langle x_i, \lambda_j x_j \rangle \quad (\because A x_j = \lambda_j x_j) \\
&= \lambda_j \langle x_i, x_j \rangle
\end{aligned}$$

$$\lambda_i \langle x_i, x_j \rangle = \lambda_j \langle x_i, x_j \rangle$$

$\hookrightarrow (\lambda_i - \lambda_j) \langle x_i, x_j \rangle = 0$

Since $\lambda_i \neq \lambda_j$, $(\lambda_i - \lambda_j) \neq 0$

$\therefore \langle x_i, x_j \rangle = 0$

$x_i^T x_j = 0$

x_i & x_j is orthogonal to each other.

Eigendecomposition of a Symmetric Matrix

- Now, the matrix with eigenvectors are actually **orthogonal**
- So, the inverse of the matrix could be replaced by the transpose which is much easier than handling an inverse.

$$\boxed{A = Q \Lambda Q^{-1} = Q \Lambda Q^{T}}$$

where $A \in \mathbb{R}^{m \times m}$

↳ Symmetric Square Matrix

$Q \in \mathbb{R}^{m \times m}$

↳ Orthogonal Matrix

If Q is orthogonal,

$Q^{-1} = Q^{T}$

$\Lambda \in \mathbb{R}^{m \times m}$

↳ Diagonal Matrix (Eigenvalues in the diagonal)

Positive Definiteness

- Positive definite matrix helps us solve optimization problems, decompose the matrix into a more simplified matrix, etc.
- To determine if the matrix is positive definite or not, we check the following conditions.

    1) check if the matrix is symmetric

    2) check if all eigenvalues are positive

    3) check if all the sub-determinants are also positive

    4) check if the quadratic form is positive

[1]Slides on Concepts to Linear Algebra are Prepared from Sho Nakagome's Blog on Medium..

Quadratic Form

- Let's define and check what's a quadratic form is.

$$x \in \mathbb{R}^{m\times 1}, \quad \mathbb{A} \in \mathbb{R}^{m\times m}$$

$$\boxed{\text{Quadratic form : } x^T \mathbb{A} x}$$

$x^T \mathbb{A} x$ is a scalar value.

$$(1\times m)\times(m\times m)\times(m\times 1) \rightarrow 1\times 1$$

Quadratic Form

$$x^T A x = [x_1, x_2, \cdots, x_m] \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$= [x_1, x_2, \cdots, x_m] \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mm}x_m \end{bmatrix}$$

$$\begin{aligned} = &\, x_1(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m) \\ &+ x_2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m) \\ &\;\;\vdots \\ &+ x_m(a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mm}x_m) \end{aligned}$$

$$= \sum_{i \le j}^{m} a_{ij}\, x_i\, x_j$$

Quadratic Form

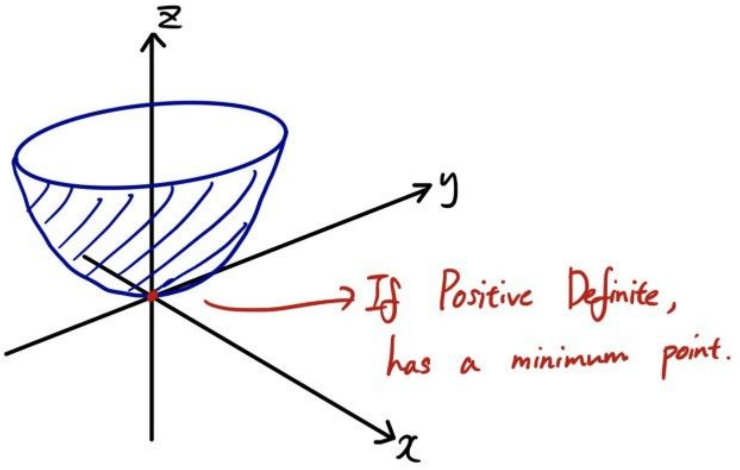
Based on the signs of the ***quadratic form***, we can classify the definiteness into three categories:

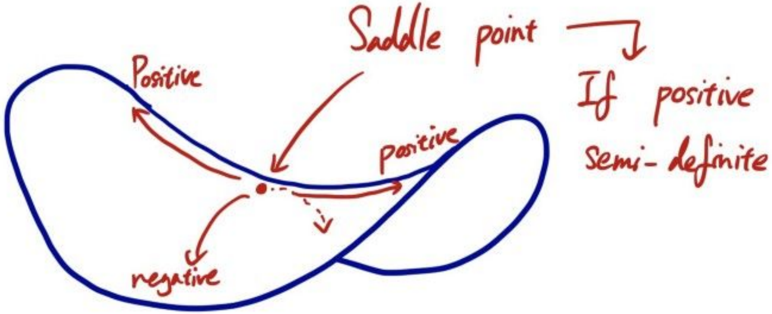
- ***Positive definite if (Quadratic form) > 0***
- ***Positive semi-definite if (Quadratic form) ≥ 0***
- ***Negative definite if (Quadratic form) < 0***

Geometric Interpretation of Positive Definiteness

- If the matrix is positive definite, then it's great because you are guaranteed to have the minimum point.

- When the matrix is positive-semidefinite, it has a somewhat stable point called a **saddle point**, but most of the time it just slips off the saddle point to keep going down to the hell where optimization becomes challenging.





Making a Positive Definite Matrix with a Matrix that's not Symmetric

- If a matrix "*A*" is not symmetric, we can still make a use of positive definiteness.
- Matrix A^TA is symmetric and square and the quadratic form of such a matrix is -

$$x^T A^T A x = (Ax)^T (Ax) = \|Ax\|^2 > 0$$

- ***So, we can simply multiply the matrix that's not symmetric by its transpose and the product will become symmetric, square, and positive definite!***

Singular Value Decomposition

- Unlike eigendecomposition where the matrix has to be a square matrix, ***SVD*** allows ***to decompose a rectangular matrix***.

$$A = U\Sigma V^T$$

$$\underset{m \times n}{\overset{\mathrm{A}}{\begin{pmatrix} x_{11} & x_{12} & & x_{1n} \\ & & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}} = \underset{m \times m}{\overset{\mathrm{U}}{\begin{pmatrix} u_{11} & & u_{m1} \\ & \ddots & \\ u_{1m} & & u_{mm} \end{pmatrix}}} \underset{m \times n}{\overset{\boldsymbol{\Sigma}}{\begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ & & & \ddots \\ 0 & & & 0 \end{pmatrix}}} \underset{n \times n}{\overset{\mathrm{V}^{\mathrm{T}}}{\begin{pmatrix} v_{11} & & v_{1n} \\ & \ddots & \\ v_{n1} & & v_{nn} \end{pmatrix}}}$$

Singular Value Decomposition

- Unlike eigendecomposition where the matrix has to be a square matrix, ***SVD*** allows ***to decompose a rectangular matrix***.

$$A = U \Sigma V^T$$

where,

$A \in R^{m \times n}$ (Real or Complex matrix)

$U \in R^{m \times m}$ (Real or Complex ***Unitary*** matrix)

$\Sigma \in R^{m \times n}$ (Real or Complex ***Diagonal*** matrix)

$V \in R^{n \times n}$ (Real or Complex ***Unitary*** matrix)

Intuitive Explanation of Singular Value Decomposition

$$\begin{cases} u_1 = A v_1 \\ u_2 = A v_2 \end{cases}$$

To make it orthonormal normalize u with singular values

$$\begin{cases} \sigma_1 u_1 = A v_1 \\ \sigma_2 u_2 = A v_2 \end{cases}$$

$$A[v_1 \ v_2]$$

$$= [u_1 \ u_2]\begin{bmatrix}\sigma_1 & 0 \\ 0 & \sigma_2\end{bmatrix} \Rightarrow \boxed{AV = U\Sigma} \Rightarrow A = U\Sigma V^{-1}$$

$$= U\Sigma V^{T}$$

Singular Value Decomposition

- Unlike eigendecomposition where the matrix has to be a square matrix, ***SVD*** allows ***to decompose a rectangular matrix***.

$$A = U\Sigma V^T$$

- So, "U" and "V" are *orthogonal* matrices when real. i.e.

$$UU^T = U^TU = I$$
$$VV^T = V^TV = I$$

Singular Value Decomposition: How to get U, $\boldsymbol{\Sigma}$, and V?

$$A = U \Sigma V^T$$

$$A^T = V \Sigma U^T$$

$$\text{(i)} \quad A^T A = V \Sigma U^T U \Sigma V^T$$

$$= V |\Sigma|^2 V^T \quad (\because U^T U = I)$$

$A^T A$ is Positive Definite!

$\therefore V$ is orthonormal eigenvectors of $A^T A$

(Recall $\underset{\text{symmetric}}{A} = Q \Lambda Q^T$)

$$\text{(ii)} \quad A A^T = U \Sigma V^T V \Sigma U^T$$

$$= U |\Sigma|^2 U^T$$

$\hookrightarrow$ Get U

Singular Value Decomposition

- "U" and "V" are orthogonal matrices with orthonormal eigenvectors chosen from AA^T and A^TA respectively.
- Σ is a diagonal matrix with r elements equal to the root of the positive eigenvalues of AA^T or A^TA (both matrices have the same positive eigenvalues anyway).

$$\begin{pmatrix} \sqrt{\lambda_1} & & & & \\ & \sqrt{\lambda_2} & & & \\ & & \ddots & & \\ & & & \sqrt{\lambda_r} & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \equiv \overset{\Sigma}{\begin{pmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix}}$$

σ_2: singular value

Factorization of User-Item Rating Matrix

- The problems of cold-start and limited coverage can also be alleviated by factoring the user-item rating matrix.

Factorization of User-Item Rating Matrix

Let ***R*** be a user-item rating matrix of rank ***n*** and the order of ***|U| x |I|***.

- We wish to approximate ***R*** with a matrix $\hat{R} = PQ^T$ of lower rank *k* < *n* by minimizing the following objective function.

$$
\begin{aligned}
E(P, Q) &= ||R - PQ^\top||_F^2 \\
&= \sum_{u,i} \left(r_{ui} - \mathbf{p}_u \mathbf{q}_i^\top\right)^2 .
\end{aligned}
$$

- Finding the factor matrices P and *Q* is equivalent to computing the ***Singular Value decomposition of R.***

$$R = P\Sigma Q^T$$

Factorization of User-Item Rating Matrix

- A user *u*'s rating of item *i*, which is denoted by r_{ui}, leading to the estimate -

$$\hat{r}_{ui} = q_i^T p_u.$$

- Applying SVD in the collaborative filtering domain requires factoring the user-item rating matrix.
- Conventional SVD is undefined when knowledge about the matrix is incomplete.
- Applying SVD on user-item rating matrix raises difficulties due to its sparsity.
- Moreover, carelessly addressing only the relatively few known entries is highly prone to overfitting.

Factorization of User-Item Rating Matrix

- Recent works suggested modeling directly the observed ratings only, while avoiding overfitting through a regularized model.
- To learn the factor vectors (p_u and q_i), the system minimizes the regularized squared error on the set of known ratings:

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

Here, κ is the set of the (u, i) pairs for which r_{ui} is known (the training set).

Next Lecture

- Collaborative Methods: Model-Based Contd…