Matrix Factorization

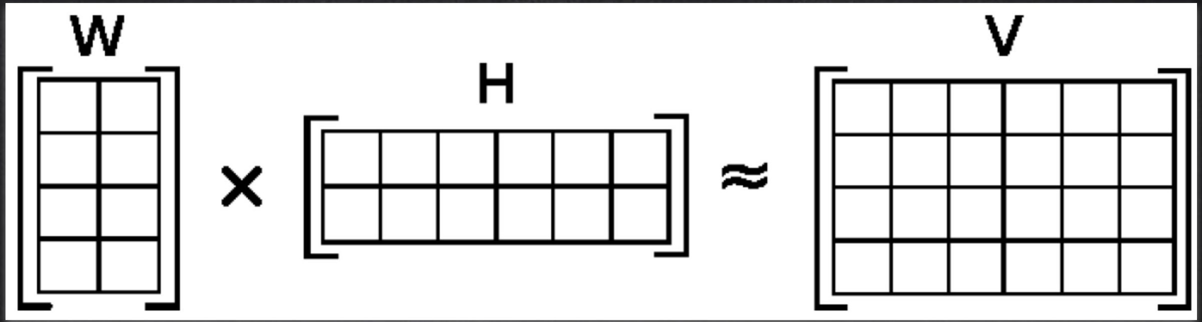
Group 8: Ellen, Circle, Joyce, Hang

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

1.

Definition of Matrix Factorization

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC



Big concept

Matrix factorization is a class of collaborative filtering algorithms used in recommender systems. Matrix factorization algorithms work by decomposing the user-item interaction matrix into the product of two lower dimensionality rectangular matrices.

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Receipt & Function of MF

- ✘ Define a model
- ✘ Define an objective function
- ✘ Optimize with SGD

- ✘ **MF serves the collaborative filtering**

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

2.

Methods of Matrix Factorization

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Choosing the objective function

Observed Only MF

1		1	1	
	1			1
1	1	1		
			1	1

$$\Sigma_{(i,\, j) \in \text{obs}} (A_{ij} - U_i \,.\, V_j)^2$$

Weighted MF

1	0	1	1	0
0	1	0	0	1
1	1	1	0	0
0	0	0	1	1

$$\Sigma_{(i,\, j) \in \text{obs}} (A_{ij} - U_i \,.\, V_j)^2 + w_0 \, \Sigma_{(i,\, j) \notin \text{obs}} (0 - U_i \,.\, V_j)^2$$

SVD

1	0	1	1	0
0	1	0	0	1
1	1	1	0	0
0	0	0	1	1

$$|A - U V^T|_F^{\,2} = \Sigma_{(i,\, j)} (A_{ij} - U_i \,.\, V_j)^2$$

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Observed only MF

$$\min_{U \in \mathbb{R}^{m\times d},\ V \in \mathbb{R}^{n\times d}} \sum_{(i,j)\in \text{obs}} (A_{ij} - \langle U_i, V_j \rangle)^2.$$

- ✘ Squared distance
- ✘ Only sum over all pairs of observed entries, that are non-zero values in the feedback matrix.
- ✘ Not a good idea – a matrix of all ones will have a minimal loss and produce a model that can't make effective recommendations and that generalizes poorly.

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Weighted Matrix Factorization

$$\min_{U\in\mathbb{R}^{m\times d},\, V\in\mathbb{R}^{n\times d}} \sum_{(i,j)\in \text{obs}} (A_{ij} - \langle U_i, V_j\rangle)^2 + w_0 \sum_{(i,j)\notin \text{obs}} (\langle U_i, V_j\rangle)^2.$$

- Weighted Matrix Factorization
- Decompose the objective into the following two sums:
  - A sum over observed entries.
  - A sum over unobserved entries (treated as zeroes).
- w_0 is a hyperparameter that weights the two terms so that the objective is not dominated by one or the other.

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Singular Value Decomposition (SVD)

$$\min_{U \in \mathbb{R}^{m \times d},\, V \in \mathbb{R}^{n \times d}} \|A - UV^T\|_F^2.$$

- ✘ Squared Frobenius distance
- ✘ Not a good idea – the matrix A may be very sparse. The solution will likely be close to zero, leading to poor generalization performance.

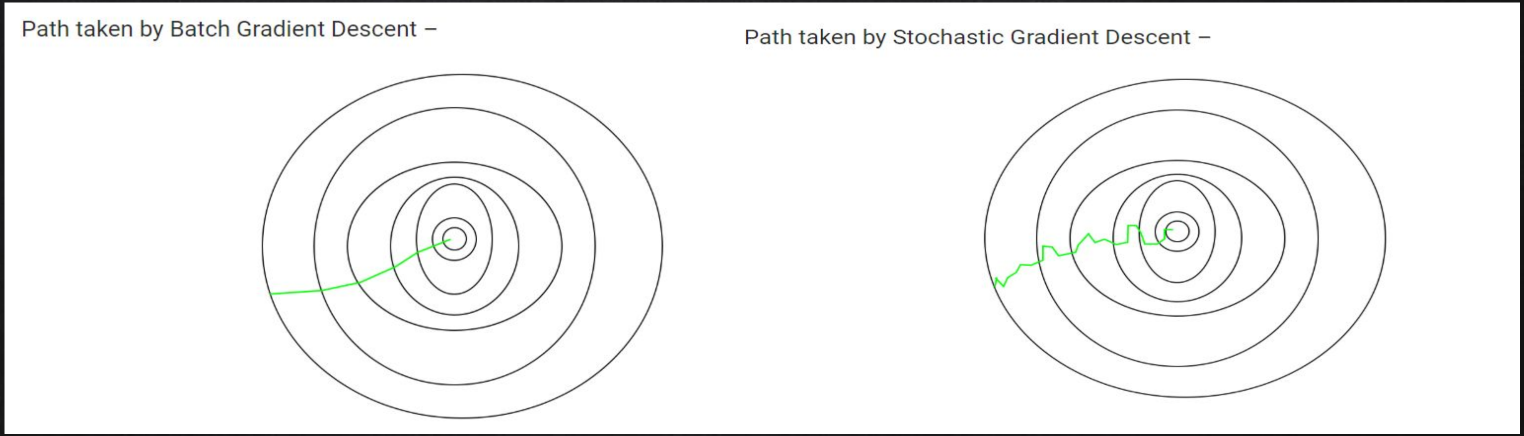
CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

3.

Methods of Optimization

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Stochastic Gradient Descent



- Randomly select a few samples instead of the whole data set
- Generally noisier than typical Gradient Descent

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Weighted Alternating Least Squares

- ✘ The objective is quadratic in each of the two matrices U and V
- ✘ Alternating between: Fixing U and solving for V
  Fixing V and solving for U

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC

Minimizing the Objective Function

- Stochastic gradient descent (SGD)
- Weighted Alternating Least Squares (WALS)

SGD	WALS
👍 Very flexible—can use other loss functions.	👎 Reliant on Loss Squares only.
👍 Can be parallelized.	👍 Can be parallelized.
👎 Slower—does not converge as quickly.	👍 Converges faster than SGD.
👎 Harder to handle the unobserved entries (n	👍 Easier to handle unobserved entries.

CONFIDENTIAL - NOT FOR USE AND/OR DISTRIBUTION TO THE GENERAL PUBLIC