User/Movie	P_1	P_2	…	P_M
U_1	3	4	… ? …	?
U_2	2	?	…	4
U_3	4	5	…	?
⋮	⋮			
U_N	5	1	… ? … ? …	3

N

User/Movie	P_1	P_2	. . . . .	.. P_M
U_1	3	4	.... ?	?
U_2	2	?	....	4
U_3	4	5	....	?
⋮	⋮			
U_N	5	1	...?... ?...	3

M

PREDICT

User/Movie	P_1	P_2	...	P_M
U_1	3	4	... ? ...	?
U_2	2	?	...	4
U_3	4	5	...	?
⋮	⋮			
U_N	5	1	... ? ... ? ...	3

N

M

Let's consider a subset of users and movies and assume complete data

	Sholay	Swades	Batman	Interstellar	Shawshank
U_1	5	5	3	3	2
U_2	2	4	5	5	3
U_3	2	2	3	3	5
⋮	.	.	.	.	.
⋮	.	.	.	.	.

Let's consider a subset of users and movies

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
U2	2	4	5	5	3
U3	2	2	3	3	5
⋮	.	.	.	.	.
⋮	.	.	.	.	.

What can you say about U1?

Let's consider a subset of users and movies

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
U2	2	4	5	5	3
U3	2	2	3	3	5
⋮	.	.	.	.	.
⋮	.	.	.	.	.

What can you say about U1? Likes Bollywood
Dislikes Hollywood

Let's consider a subset of users and movies

	Sholay	Swades	Batman	Interstellar	Shawshank
U_1	5	5	3	3	2
U_2	2	4	5	5	3
U_3	2	2	3	3	5
⋮	.	.	.	.	.
⋮	.	.	.	.	.

What can you say about U2?

Let's consider a subset of users and movies

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
U2	2	4	5	5	3
U3	2	2	3	3	5
⋮	.	.	.	.	.
⋮	.	.	.	.	.

What can you say about U2? likes "engineering"

Let's consider a subset of users and movies

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
U2	2	4	5	5	3
U3	2	2	3	3	5
⋮	.	.	.	.	.
⋮	.	.	.	.	.

What can you say about U3?

Let's consider a subset of users and movies

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
U2	2	4	5	5	3
U3	2	2	3	3	5
⋮					

What can you say about U3? Likes "shorter" movies

	Sholay	Swades	Batman	Interstellar	Shawshank
Bollywoodness	1.2	0.95	0.01	0.01	0.02
Engineering	0.1	0.8	0.9	0.95	0.01
Length	0.1	0.08	0.2	0.25	0.90

	Sholay	Swades	Batman	Interstellar	Shawshank
Bollywoodness	1.2	0.95	0.01	0.01	0.02
Engineering	0.1	0.8	0.9	0.95	0.01
Length	0.1	0.08	0.2	0.25	0.90

- Describe each movie with some "r" features
- we have created a matrix H of size $r \times M$

  #movies

	Bollywoodness	Engineering	Length
U_1	4.0	0.7	0.7
U_2	...	...	...
U_3	...	...	...

- Describe each user with some "r" features

	Bollywoodness	Engineering	Length
U_1	4.0	0.7	0.7
U_2	...	...	...
U_3	...	...	...
$\vdots$			

* Describe each user with some "r" features
* Create a matrix W of size $N_{users} \times r$ features

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2

U1 has rated sholay 5

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
	..	..	..	..	..

U1 has rated sholay 5

	Bollywoodness	Engineering	Length
U1	4.0	0.7	0.7
⋮	⋮	⋮	⋮

	Sholay	Swades	Batman
Bollywoodness	1.2	0.95	0.01
Engineering	0.1	0.8	0.9
Length	0.1	0.08	0.2

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
	..	..	..	..	..

U1 has rated sholay 5

	Bollywoodness	Engineering	Length
U1	4.0	0.7	0.7
⋮	⋮	⋮	⋮

	Sholay	Swades	Batman
Bollywoodness	1.2	0.95	0.01
Engineering	0.1	0.8	0.9
Length	0.1	0.08	0.2

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
..	..	..	..	..	..

U1 has rated sholay 5

	Bollywoodness	Engineering	Length
U1	4.0	0.7	0.7
⋮	⋮	⋮	⋮

	Sholay	Swades	Batman
Bollywoodness	1.2	0.95	0.01
Engineering	0.1	0.8	0.9
Length	0.1	0.08	0.2

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
⋮	⋮	⋮	⋮	⋮	⋮

U1 has rated Sholay 5

	Bollywoodness	Engineering	Length
U1	4.0	0.7	0.7
⋮	⋮	⋮	⋮

	Sholay	Swades	Batman
Bollywoodness	1.2	0.95	0.01
Engineering	0.1	0.8	0.9
Length	0.1	0.08	0.2

	Sholay	Swades	Batman	Interstellar	Shawshank
U1	5	5	3	3	2
	..	..	..	..	..

U1 has rated Sholay 5

	Bollywoodness	Engineering	Length
U1	4.0	0.7	0.7
⋮	⋮	⋮	⋮

	Sholay	Swades	Batman
Bollywoodness	1.2	0.95	0.01
Engineering	0.1	0.8	0.9
Length	0.1	0.08	0.2

$$4 \times 1.2 + .7 \times .1 + .7 \times .1 \approx 5$$

MATRIX FACTORISATION

$$A_{N \times M} \approx W_{N \times r} H_{r \times M}$$

* $r << N$ and $r << M$
* Also called low rank decomposition

MATRIX FACTORISATION

$$A_{N \times M} \approx W_{N \times r} \, H_{r \times M}$$

* Goal: given A, learn W and H s.t.

$$A \approx WH$$

* or;

$$W^*, H^* = \underset{W, H}{\operatorname{argmin}} \|A - WH\|_F^2$$

Aside : NORMS

$$y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

$$\|y\|_2^2 = ?$$

Aside : NORMS

$$y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

$$\|y\|_2^2 = ?$$

Above is square of ℓ_2 norm of y

$$= 1^2 + 2^2 + 3^2 + 4^2 = 30$$

Aside : NORMS

$$A = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\|A\|_F^2 = ?$$

Aside : NORMS

$$A = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 2 & 1 & 2 & 2 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$\|A\|_F^2 = ?$$

Frobenius norm

$$= \sum_{i=1}^{3} \sum_{j=1}^{4} a_{ij}^2 = 1^2 + 1^2 + 2^2 + 2^2$$
$$+ 2^2 + \dots \quad .$$
$$+ \dots \quad + 1^2$$

MATRIX FACTORISATION

$$W^*, H^* = \underset{W, H}{\operatorname{argmin}} \| A - WH \|_F^2$$

Q: How to learn W and H

MATRIX FACTORISATION

$$W^*, H^* = \underset{W, H}{\operatorname{argmin}} \| A - WH \|_F^2$$

METHOD I (Gradient Descent)

1) INIT W and H as $N \times r$ and $r \times N$ matrices

2) FOR i in $[1, \ldots ITER]$:

$$W = W - \alpha \frac{\partial \| A - WH \|_F^2}{\partial W}$$

$$H = H - \alpha \frac{\partial \| A - WH \|_F^2}{\partial H}$$

MATRIX FACTORISATION

$$W^*, H^* = \underset{W, H}{\text{argmin}} \|A - WH\|_F^2$$

METHOD I (Gradient Descent)

1) INIT W and H as $N \times r$ and $r \times N$ matrices

2) FOR i in $[1, \ldots ITER]$:

$$W = W - \alpha \frac{\partial \|A - WH\|_F^2}{\partial W}$$

$$H = H - \alpha \frac{\partial \|A - WH\|_F^2}{\partial H}$$

METHOD II (Alternating least squares)

1) INIT W
2) Till convergence
   - Fix W and learn H via least sq.
   - Fix H and learn W via least sq.

Alternating least squares (Intro, Rest in assignment)

$$A = \begin{bmatrix} \quad \end{bmatrix}_{N \times M} \qquad W = \begin{bmatrix} \quad \end{bmatrix}_{N \times r} \qquad \begin{array}{c} H \\ \begin{bmatrix} \quad \end{bmatrix}_{r \times M} \end{array}$$

Alternating least squares (Intro, Rest in assignment)

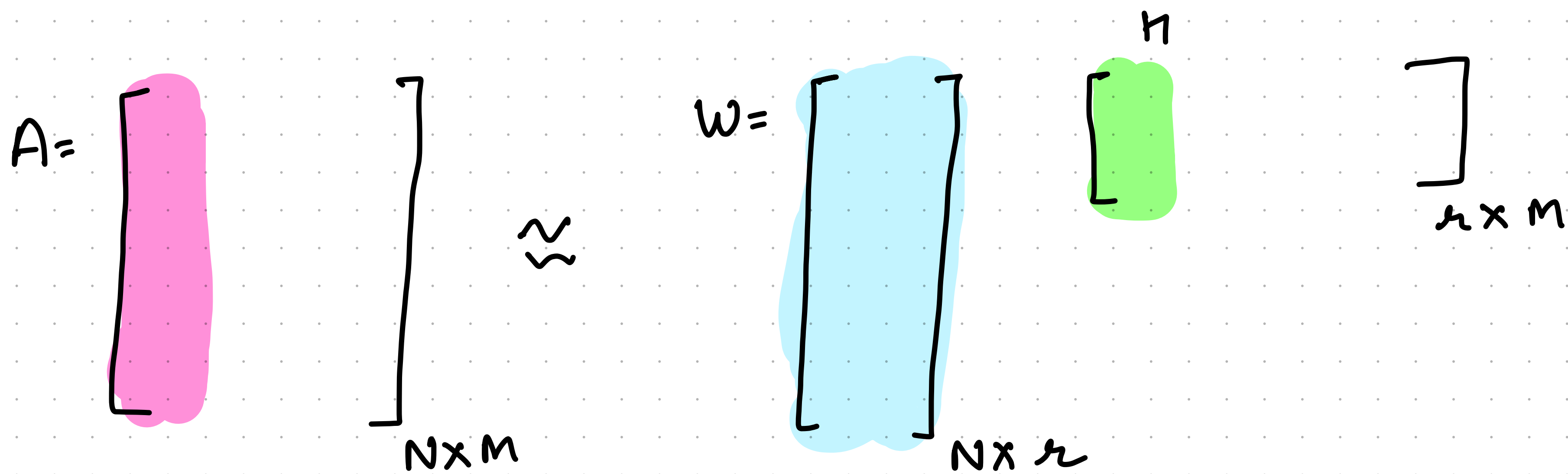
$$A = \Big[\quad \Big]_{N \times M} \approx W = \Big[\quad \Big]_{N \times r} \overset{H}{\Big[\quad \Big]}_{r \times M}$$

Remember linear regression

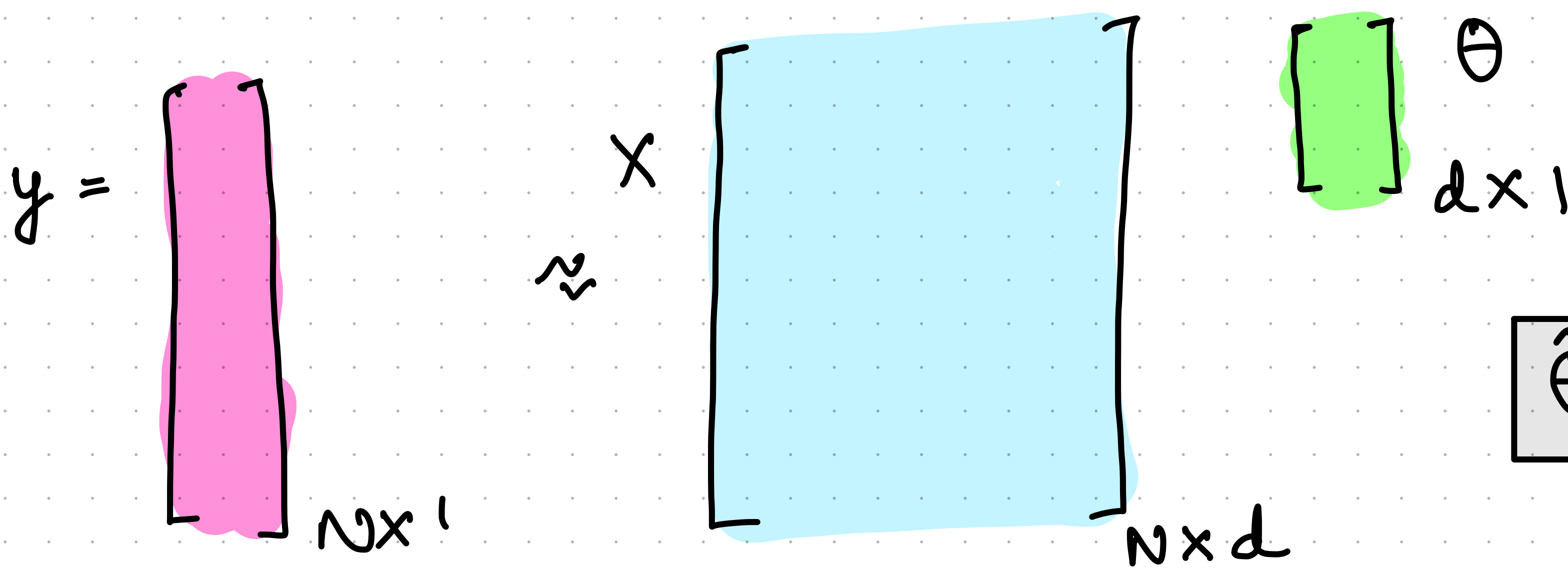
$$y = \Big[\quad \Big]_{N \times 1} \approx X \Big[\quad \Big]_{N \times d} \Big[\quad \Big]_{d \times 1} \theta$$

$$\boxed{\hat{\theta} = LS(X, y)}$$

Alternating least squares (Intro, Rest in assignment)



Remember linear regression



$$\boxed{\hat{\theta} = LS(X, y)}$$

Alternating least squares (Intro, Rest in assignment)

$$A = \left[\quad\right]_{N\times M} \approx W = \left[\quad\right]_{N\times r} \; \overset{H}{\left[\quad\right]}_{r\times M}$$

$$\boxed{\hat{H}[:,0] = LS(W, A[:,0])}$$

Remember linear regression

$$y = \left[\quad\right]_{N\times 1} \approx X \left[\quad\right]_{N\times d} \; \overset{\theta}{\left[\quad\right]}_{d\times 1}$$

$$\boxed{\hat{\theta} = LS(X, y)}$$