Matrix Factorization

April 16, 2019

Data Science CSCI 1951A

Brown University

Instructor: Ellie Pavlick

HTAs: Wennie Zhang, Maulik Dang, Gurnaaz Kaur

Announcements

- …

Today

- Dimensionality Reduction
- Matrix Factorization with SVD
- Applications to: Topic Modeling, Recommendation Systems

What is dimensionality reduction?

Clicks	Recency	Reading Level	Photo	Title: “new”	Title: “tax”	Title: “this”	…
10	1.3	11	1	1	0	0	…
1000	1.7	3	1	0	0	1	…
1000000	2.4	2	1	0	0	1	…
1	5.9	19	0	0	0	0	…

What is dimensionality reduction?

Clicks	Recency	Reading Level	Photo	Title: “new”	Title: “tax”	Title: “this”	…
10	1.3	11	1	1	0	0	…
1000	1.7	3	1	0	0	1	…
1000000	2.4	2	1	0	0	1	…
1	5.9	19	0	0	0	0	…

often 1000s or (100s of 1000s) of features

What is dimensionality reduction?

Clicks	Recency	Reading Level	Photo	Title: "new"	Title: "tax"	Title: "this"	…
10	1.3	11	1	1	0	0	…
1000	1.7	3	1	0	0	1	…
1000000	2.4	2	1	0	0	1	…
1	5.9	19	0	0	0	0	…

many (most) are redundant or useless

What is dimensionality reduction?

Clicks	Recency	Reading Level	Photo	Title: “new”	Title: “tax”	Title: “this”	…
10	1.3	11	1	1	0	0	…
1000	1.7	3	1	0	0	1	…
1000000	2.4	2	1	0	0	1	…
1	5.9	19	0	0	0	0	…

- slower to train
- easier to overfit
- harder to visualize/interpret (*)

Rank of a matrix

2	1	1
4	3	1
2	0	2
8	4	4

Rank of a matrix

2	1	1
4	3	1
2	0	2
8	4	4

Rank of a matrix

2 =	1 +	1
4 =	3 +	1
2 =	0 +	2
8 =	4 +	4

Rank = 2

Rank of a matrix

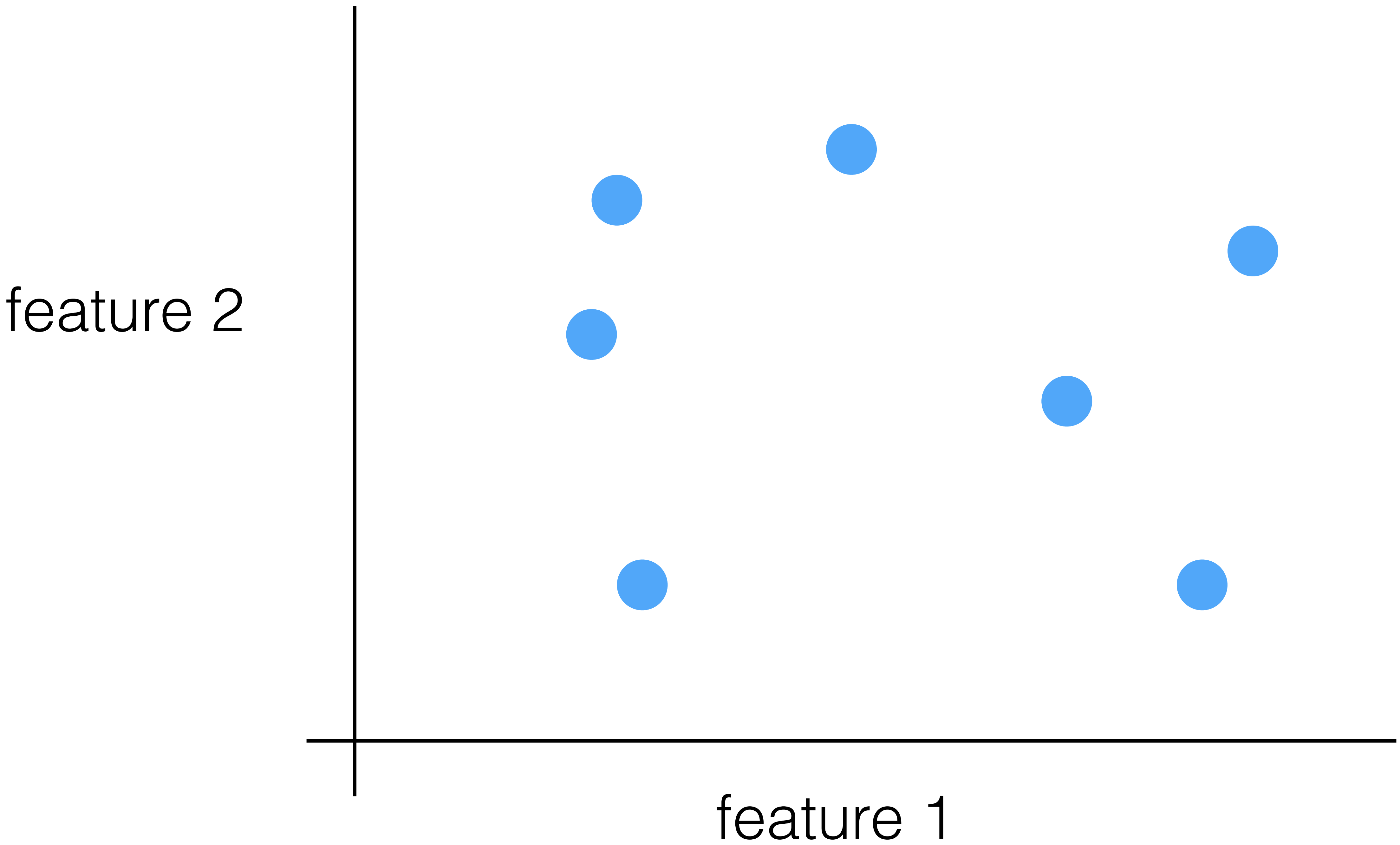
No new signal

2	1	1
4	3	1
2	0	2
8	4	4

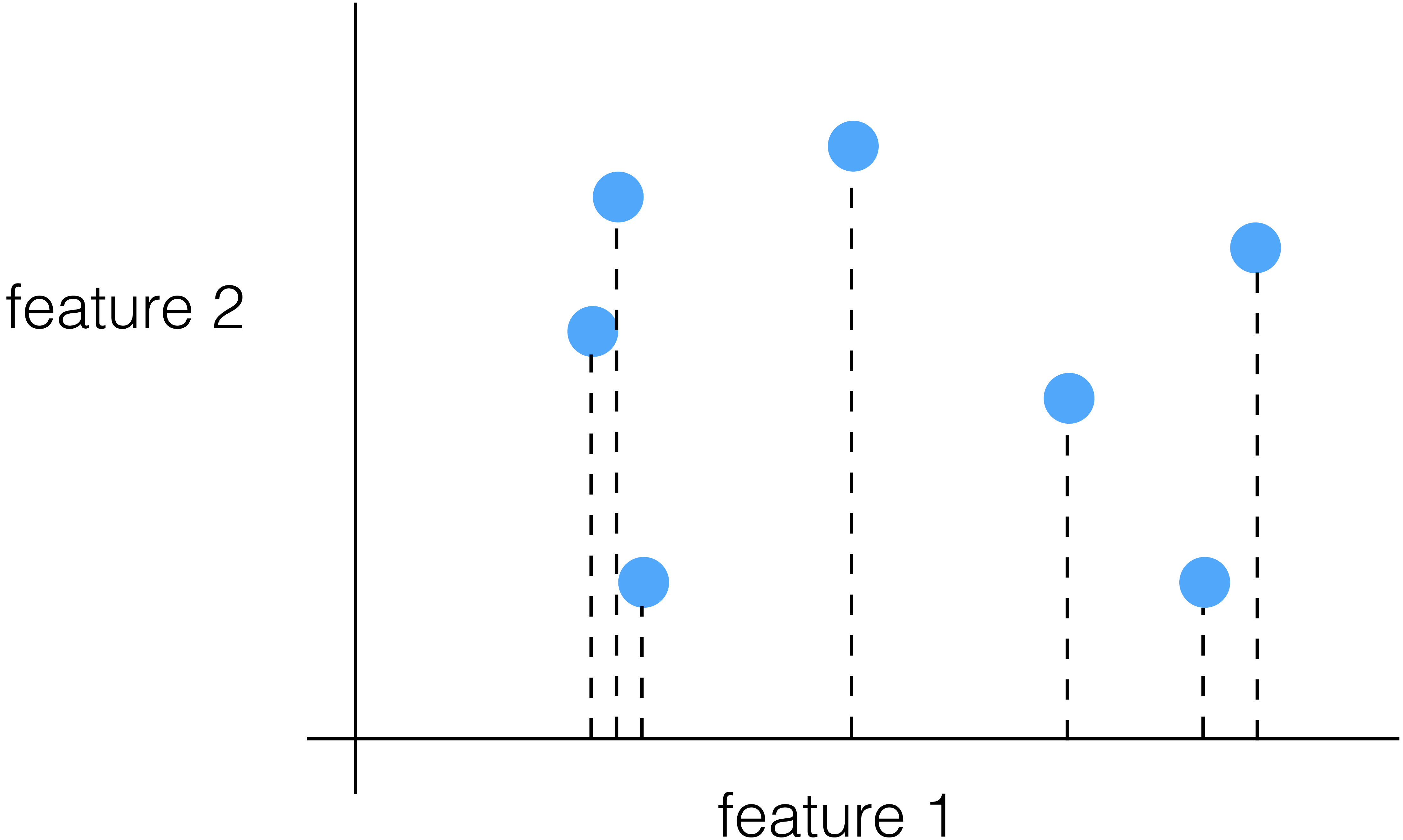
Rank = 2

Clicker Question!

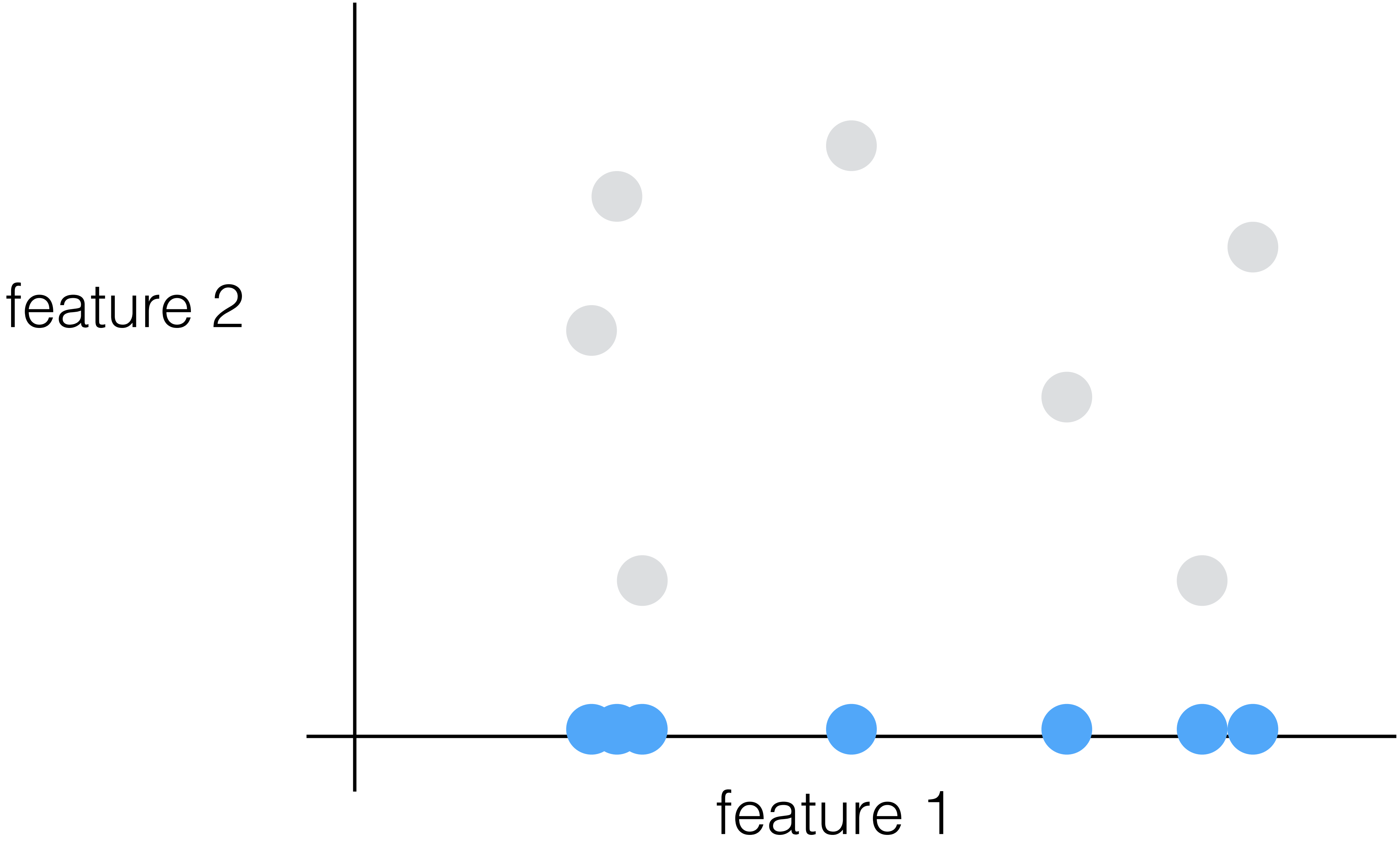
Principle Component Analysis (PCA)



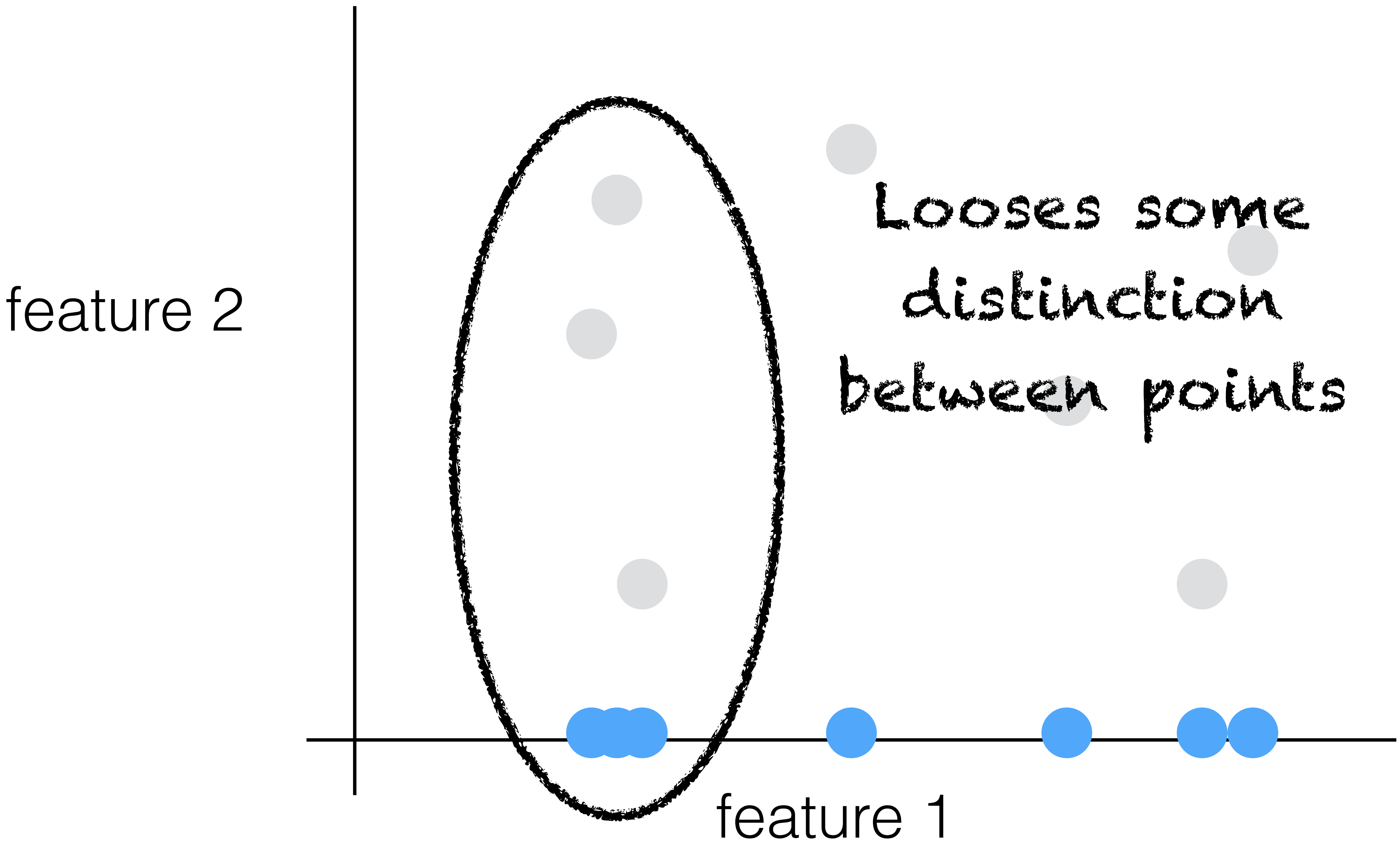
Principle Component Analysis (PCA)



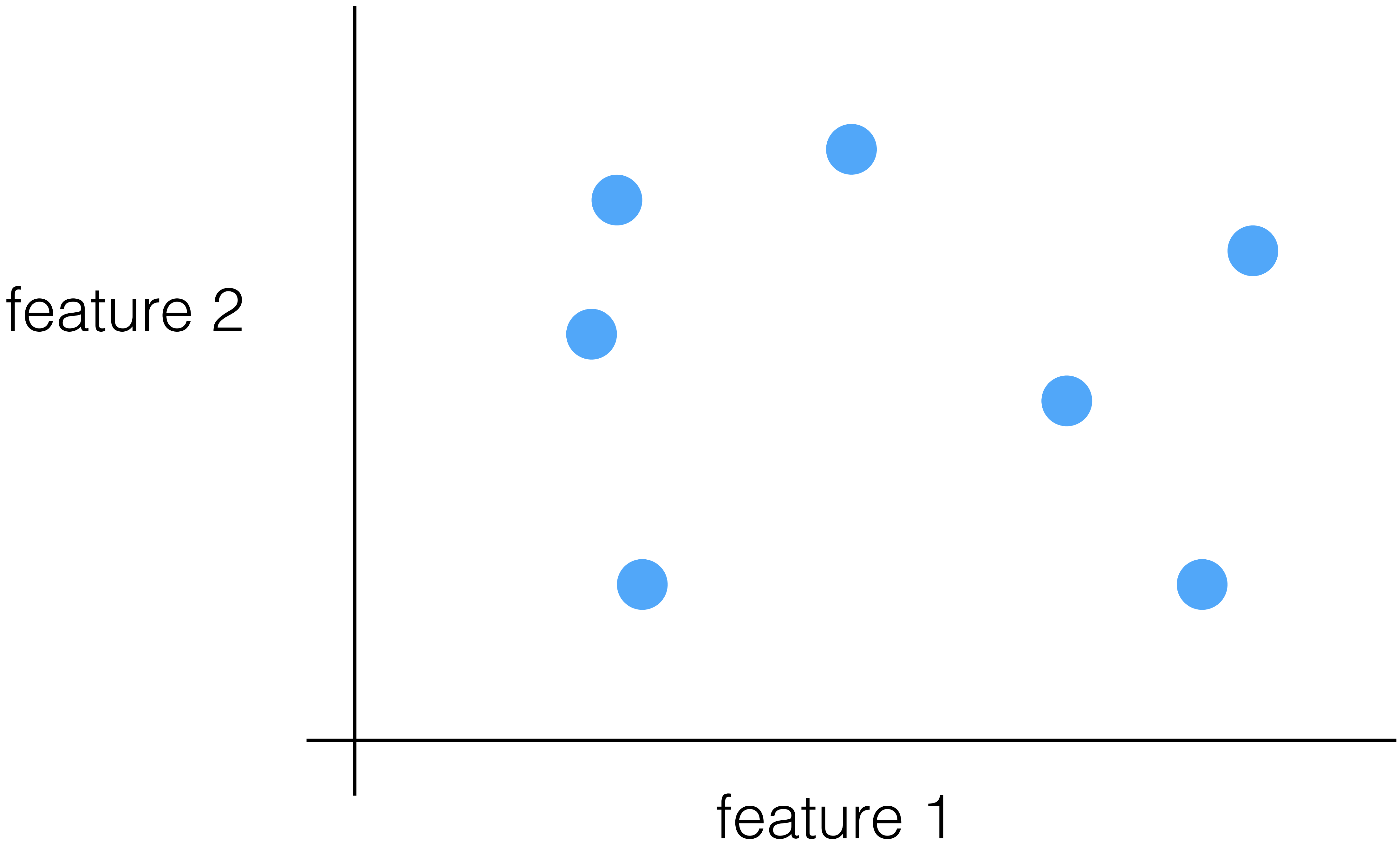
Principle Component Analysis (PCA)



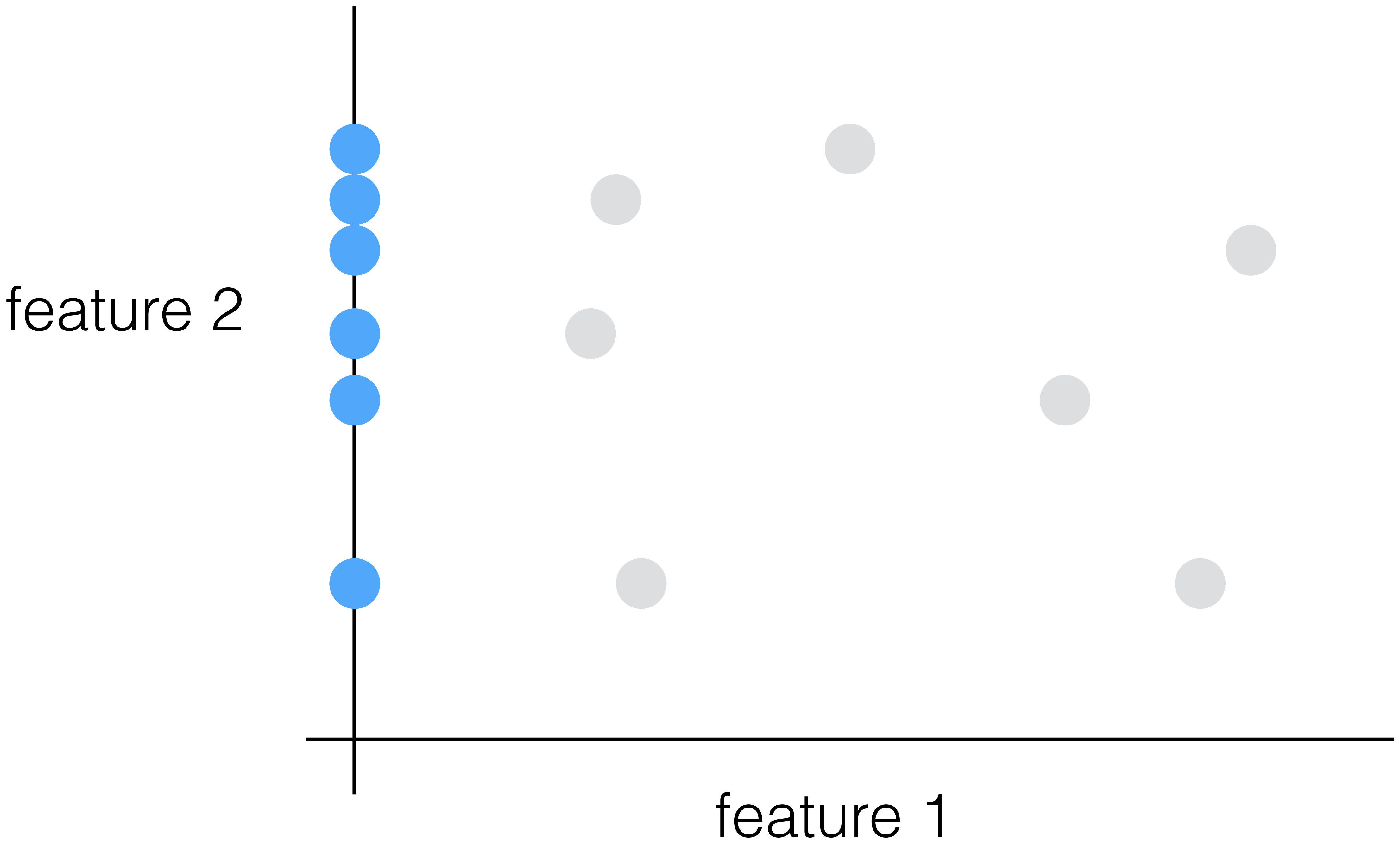
Principle Component Analysis (PCA)



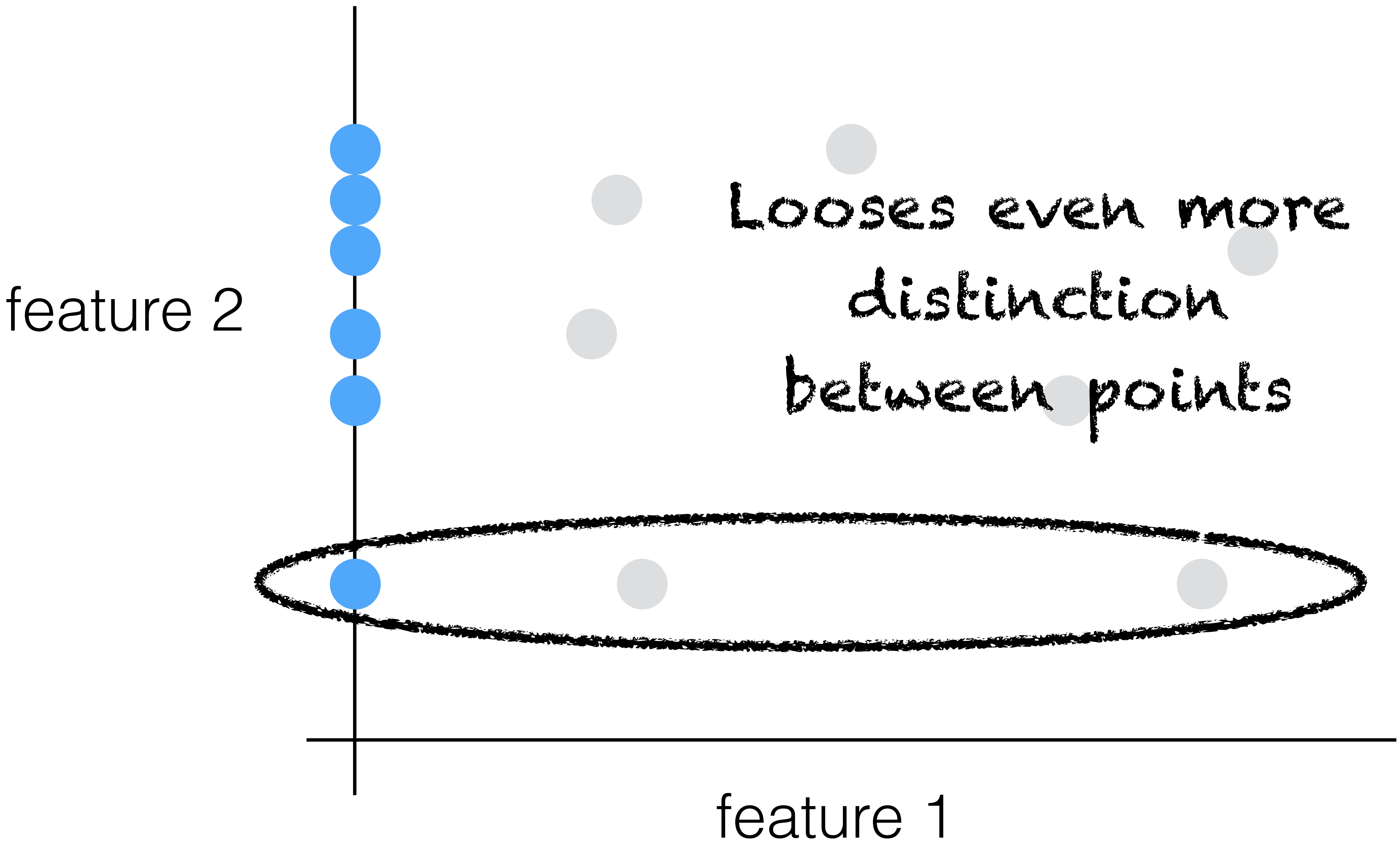
Principle Component Analysis (PCA)



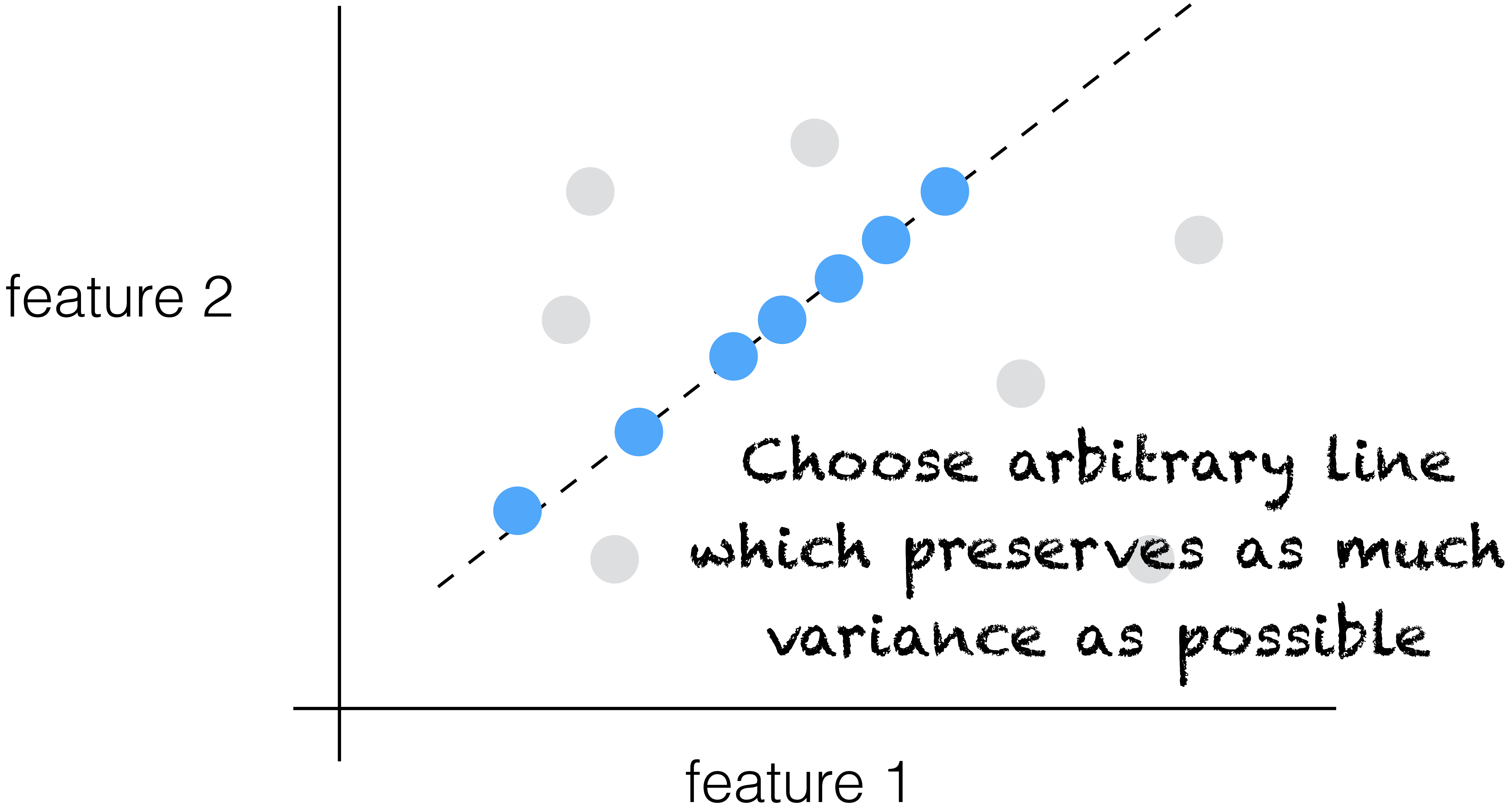
Principle Component Analysis (PCA)



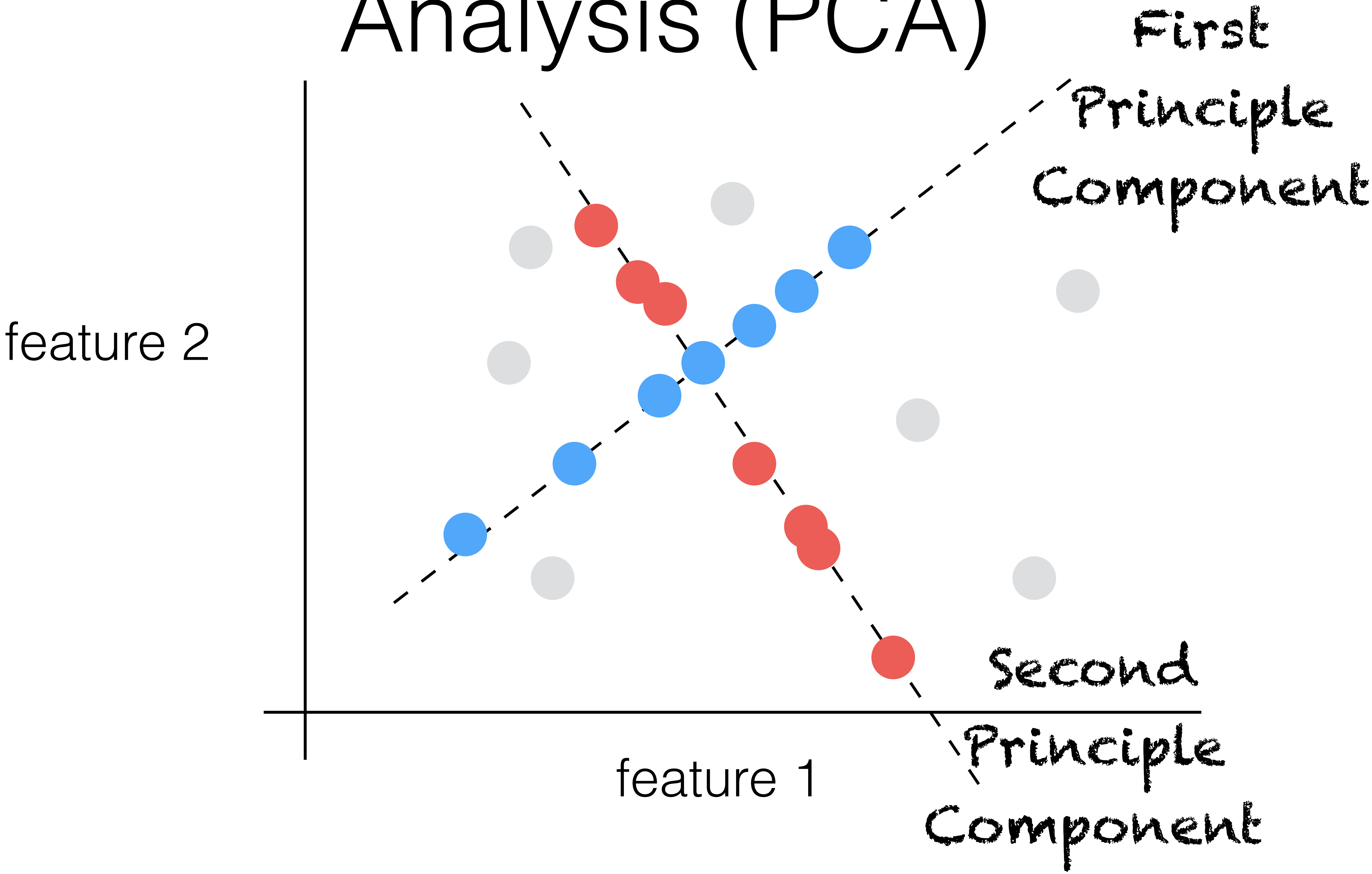
Principle Component Analysis (PCA)



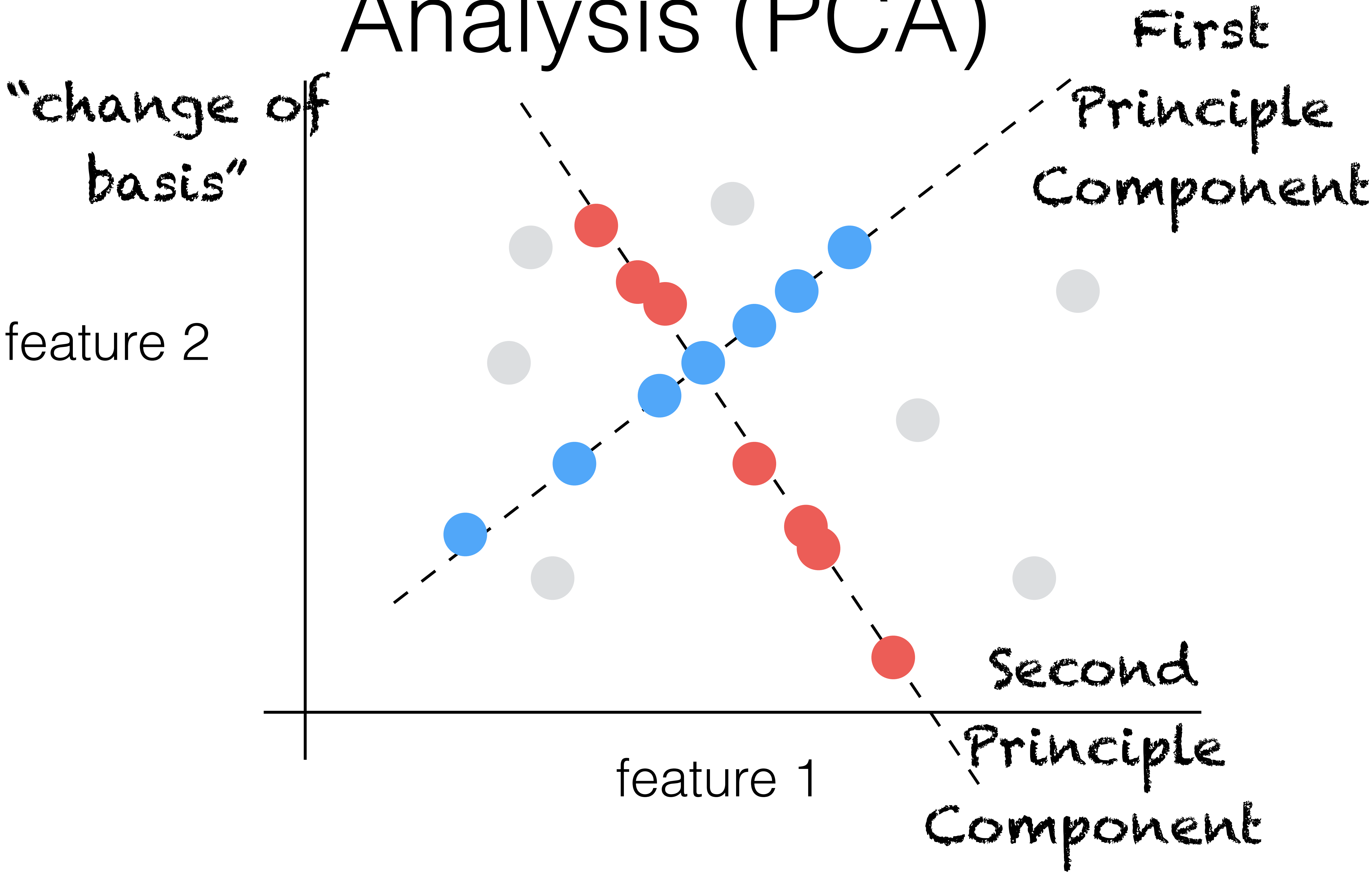
Principle Component Analysis (PCA)



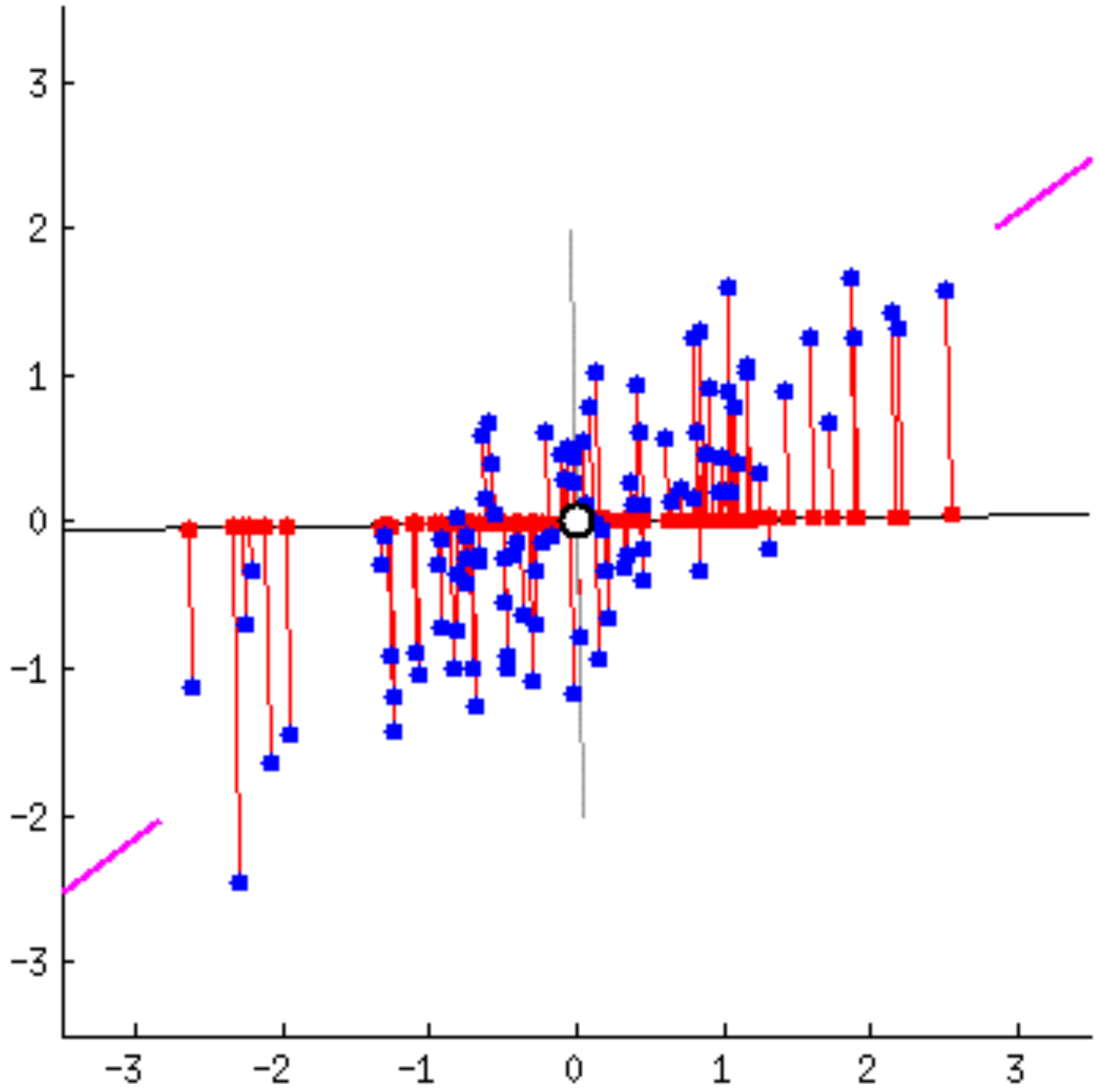
Principle Component Analysis (PCA)



Principle Component Analysis (PCA)



Principle Component Analysis (PCA)



Principle Component Analysis (PCA)

- Eigenvalue Decomposition of covariance matrix
- Singular Value Decomposition of the data matrix

Principle Component Analysis (PCA)

- Eigenvalue Decomposition of covariance matrix
- **Singular Value Decomposition of the data matrix**

Principle Component Analysis (PCA)

- Eigenvalue Decomposition of covariance matrix
- **Singular Value Decomposition of the data matrix**

Technically, PCA != SVD

Principle Component Analysis (PCA)

- Eigenvalue Decomposition of covariance matrix
- **Singular Value Decomposition of the data matrix**

Technically, PCA != SVD
(but in practice these are used interchangeably)

Matrix Arithmetic Refresh

a1	a2	a3

$$\vec{a}$$

b1	b2	b3

$$\vec{b}$$

Matrix Arithmetic Refresh

a1	a2	a3

$\vec{a}$

b1	b2	b3

$\vec{b}$

$$\vec{a} \cdot \vec{b} = (a1 \times b1) + (a2 \times b2) + (a3 \times b3)$$

Matrix Arithmetic Refresh

a11	a12	a13
a21	a22	a23
a31	a32	a33

A

b11	b12
b21	b22
b31	b32

B

Matrix Arithmetic Refresh

a11	a12	a13
a21	a22	a23
a31	a32	a33

A

3x3

b11	b12
b21	b22
b31	b32

B

3x2

Matrix Arithmetic Refresh

a11	a12	a13
a21	a22	a23
a31	a32	a33

A

3x3

b11	b12
b21	b22
b31	b32

B

3x2

Matrix Arithmetic Refresh

a11	a12	a13
a21	a22	a23
a31	a32	a33

A

3x3

b11	b12
b21	b22
b31	b32

B

3x2

??	??
??	??
??	??

AB

3x2

Matrix Arithmetic Refresh

a11	a12	a13
a21	a22	a23
a31	a32	a33

A

$m \times k$

b11	b12
b21	b22
b31	b32

B

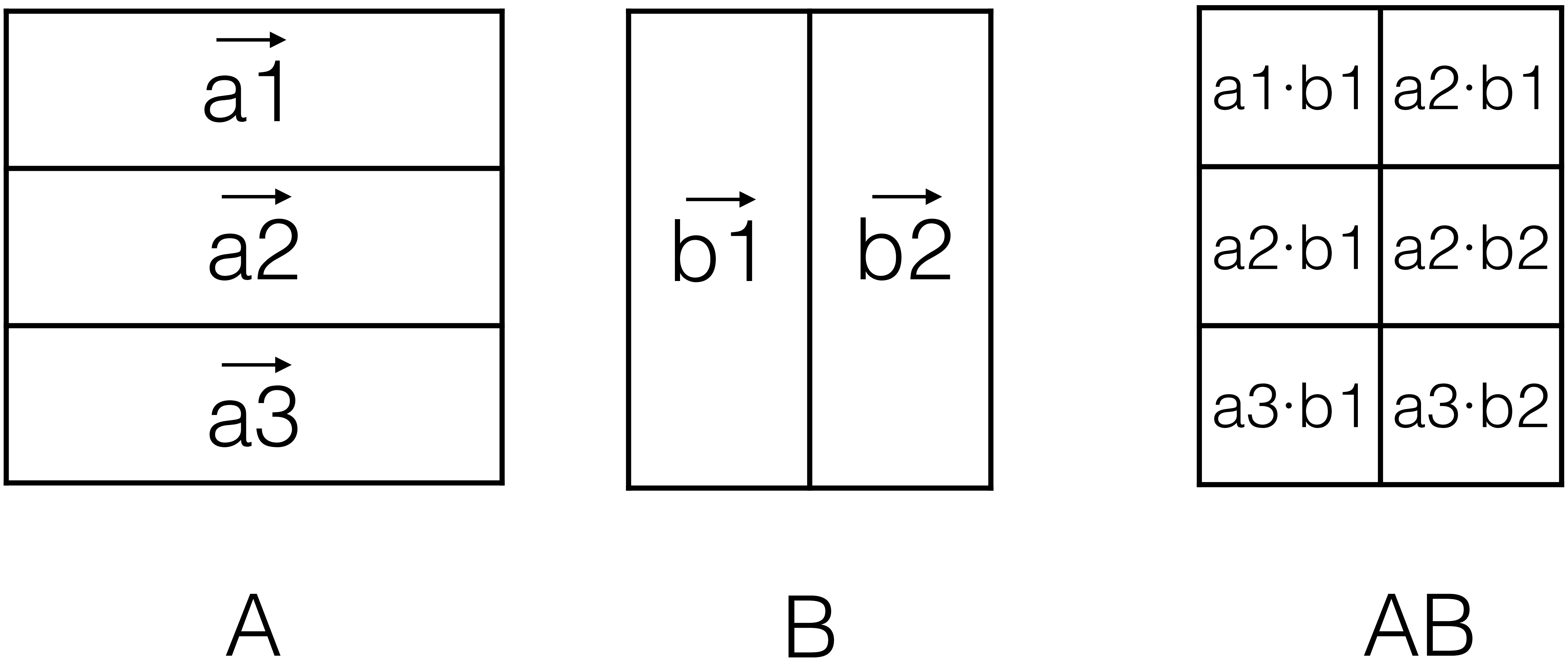
$k \times n$

??	??
??	??
??	??

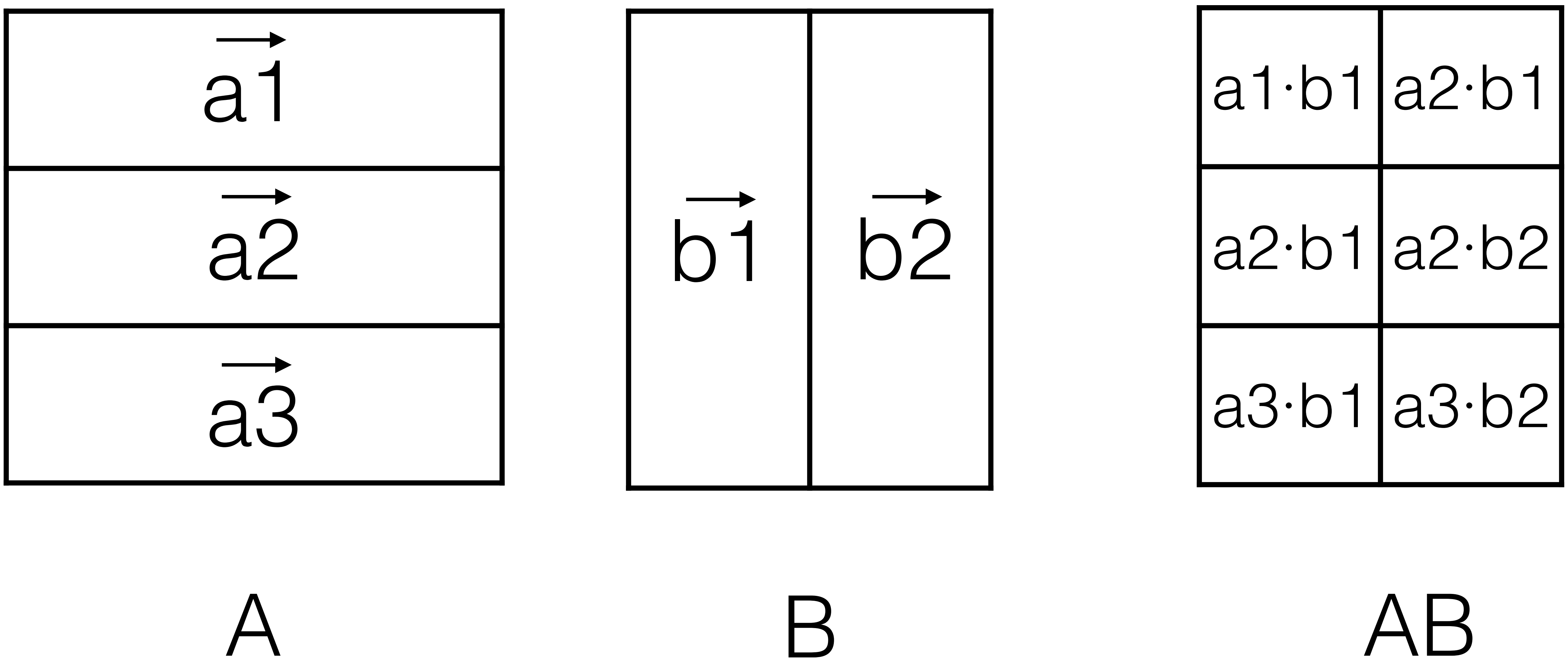
AB

$m \times n$

Matrix Arithmetic Refresh



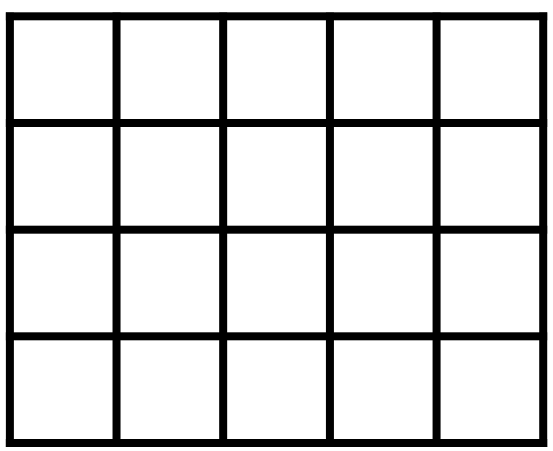
Matrix Arithmetic Refresh



AB[i][j] =ai·bj

Clicker Question!

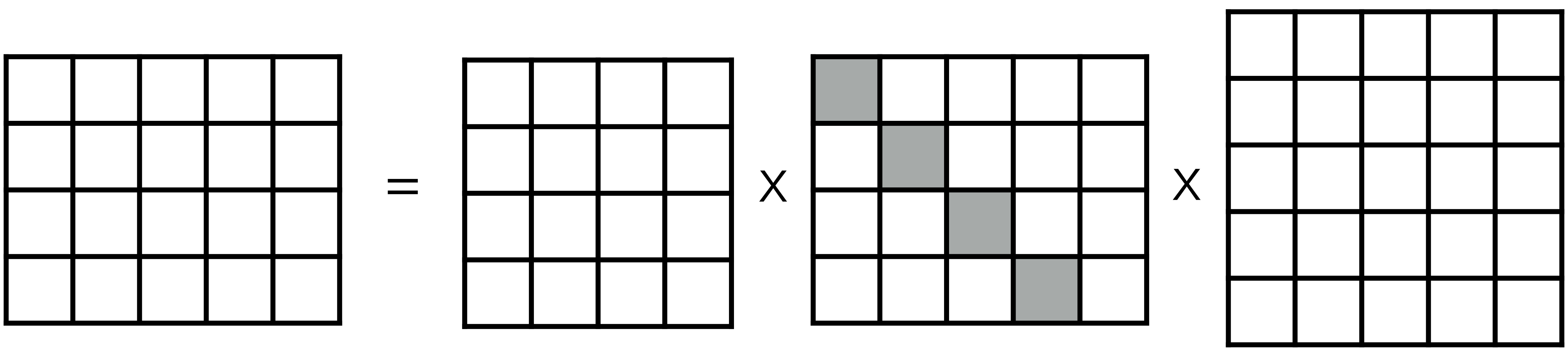
Singular Value Decomposition (SVD)



M

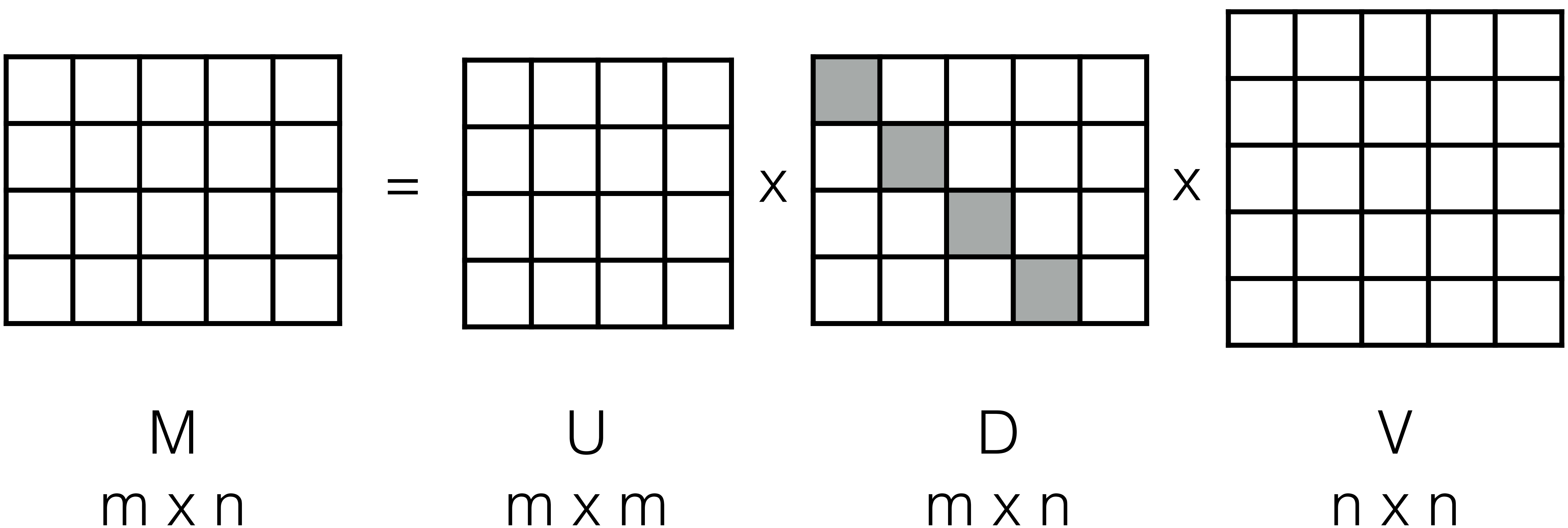
m x n

Singular Value Decomposition (SVD)



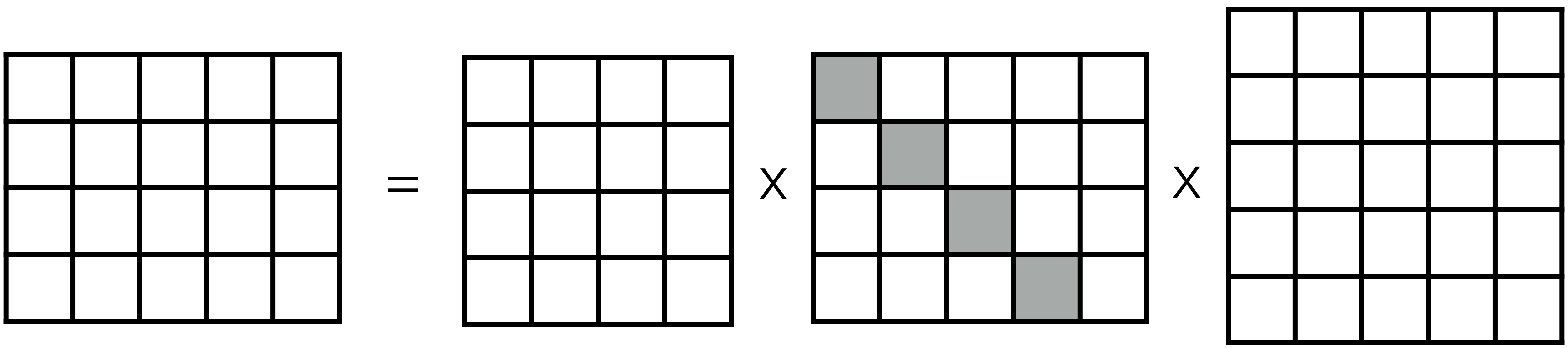
M m x n | U m x m | D m x n | V n x n

Singular Value Decomposition (SVD)



Data Matrix

Singular Value Decomposition (SVD)

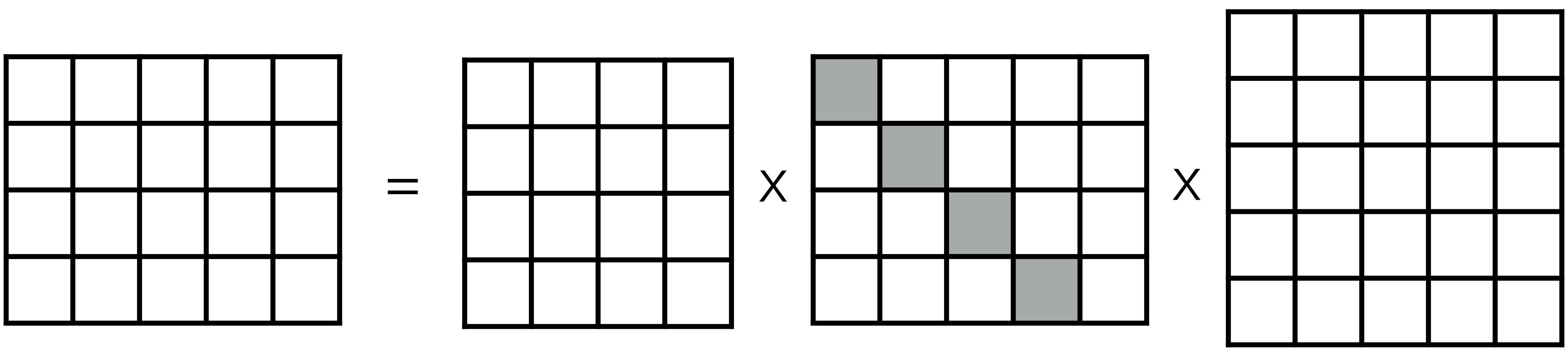


M	U	D	V
m x n	m x m	m x n	n x n

Data Matrix

Singular Values of M

Singular Value Decomposition (SVD)



M — m x n — Data Matrix

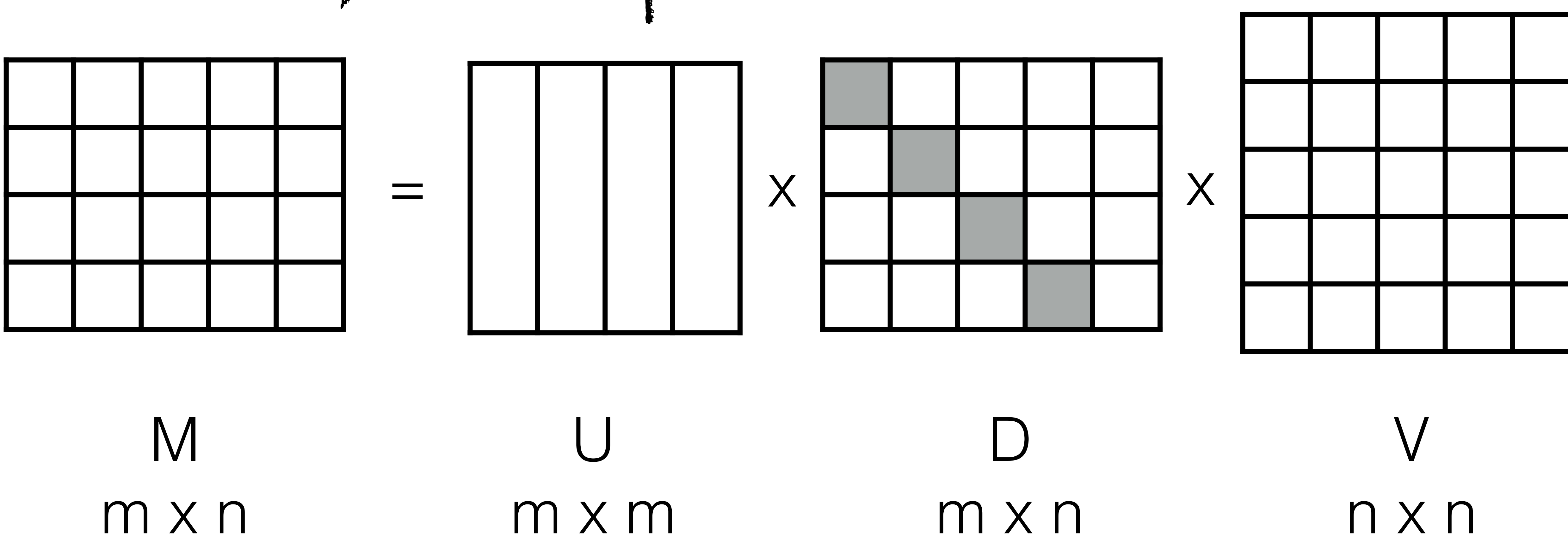
U — m x m

D — m x n — Singular Values of M (#non-zero = rank M)

V — n x n

Singular Value
on (SVD)

Representation of rows of M in new feature space



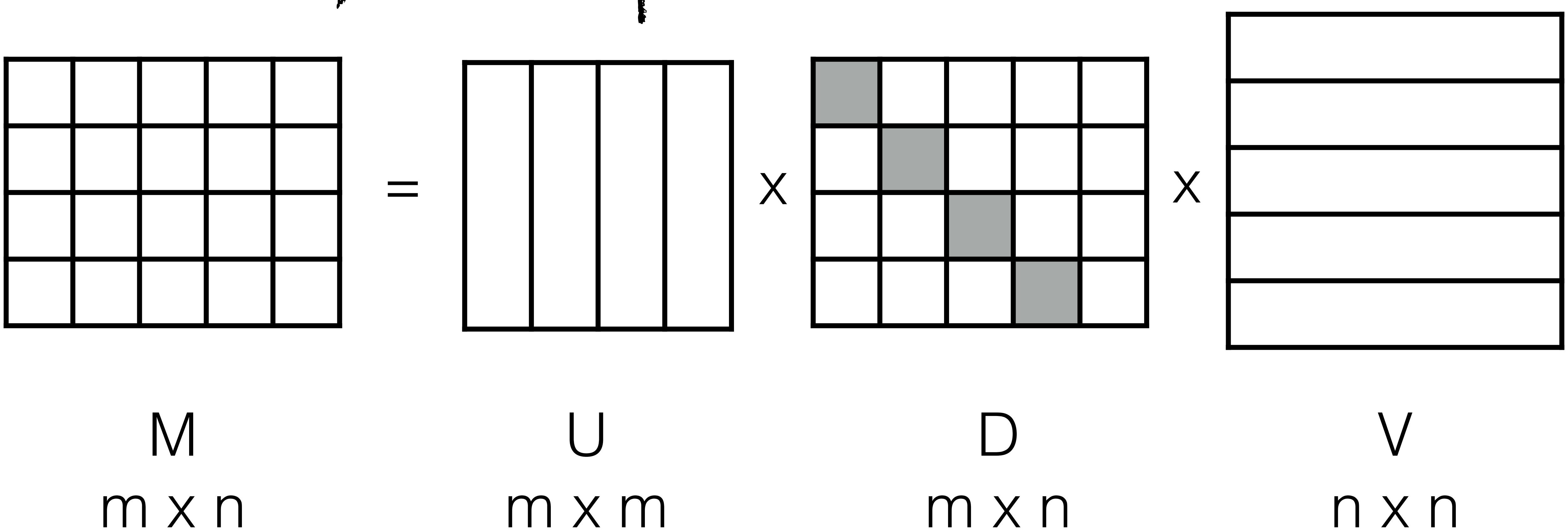
Data Matrix

Singular Values of M
(#non-zero = rank M)

Singular Value
on

Representation of rows of M in new feature space

Principle Components (new features)



Data Matrix

Singular Values of M (#non-zero = rank M)

Singular Value Decomposition (SVD)

	the	congr ess	parlia ment	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

...gular Value ...position (SVD)

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

U

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

D

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

V

gular Value
position (SVD)

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

doc1 in old feature space

U

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

D

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

V

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

gular Value oosition (SVD)

doc1 in new feature space

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

U

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

D

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

V

...gular Value ...position (SVD)

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

weight of component 1 for doc 1

U

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

D

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

V

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

[...]gular Value [...]position (SVD)

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

weight of component 1 over all the data

U

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

D

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

V

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

gular Value position (SVD)

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

component 1

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

U

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

D

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

V

gular Value
position (SVD)

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

contribution of "the" to component 1

U

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

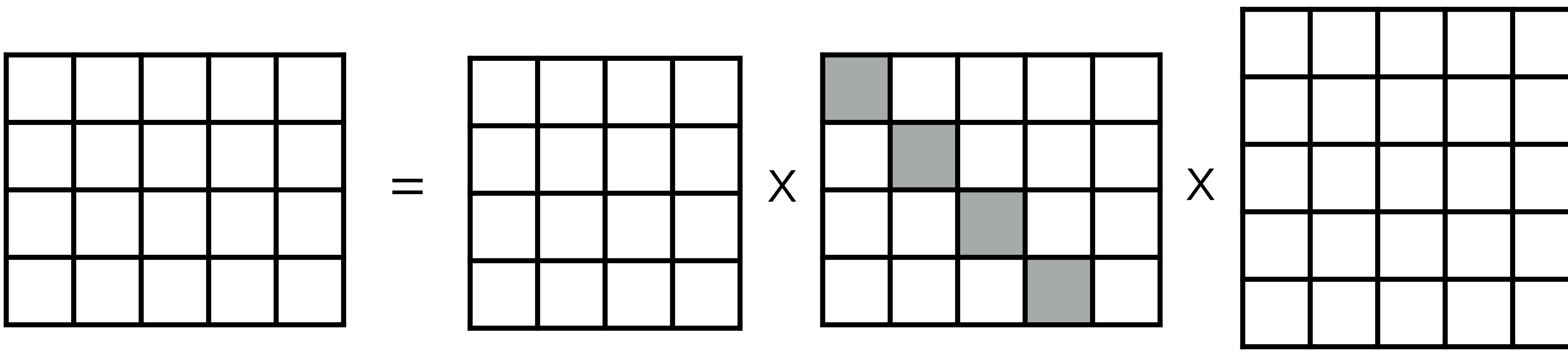
D

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

V

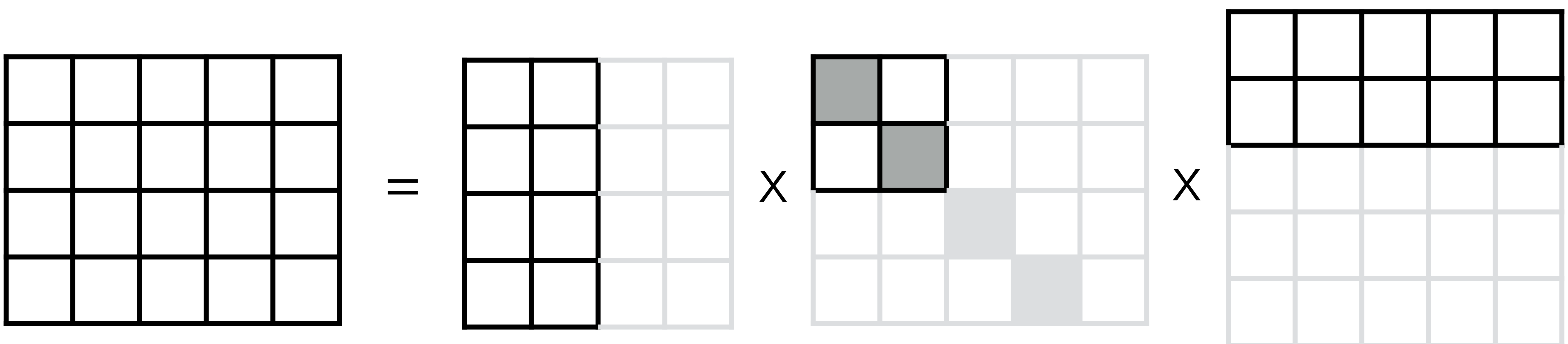
the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

Singular Value Decomposition (SVD)



M m x n | U m x m | D m x n | V n x n

Truncated Singular Value Decomposition (SVD)



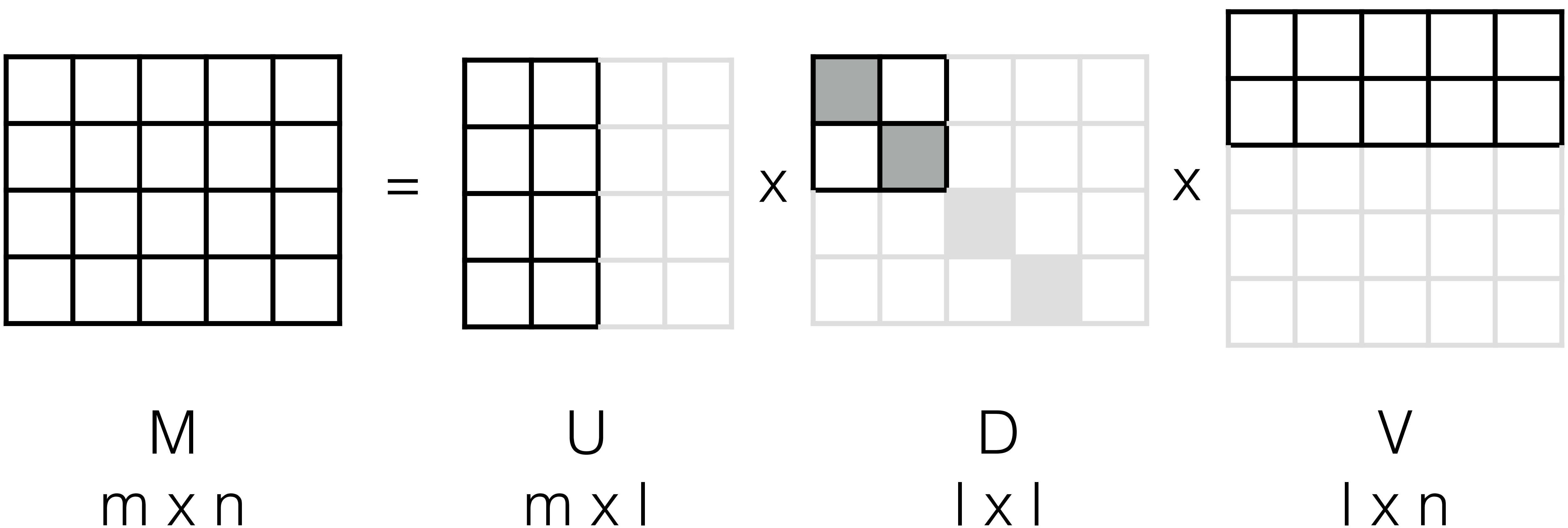
M
m x n

U
m x l

D
l x l

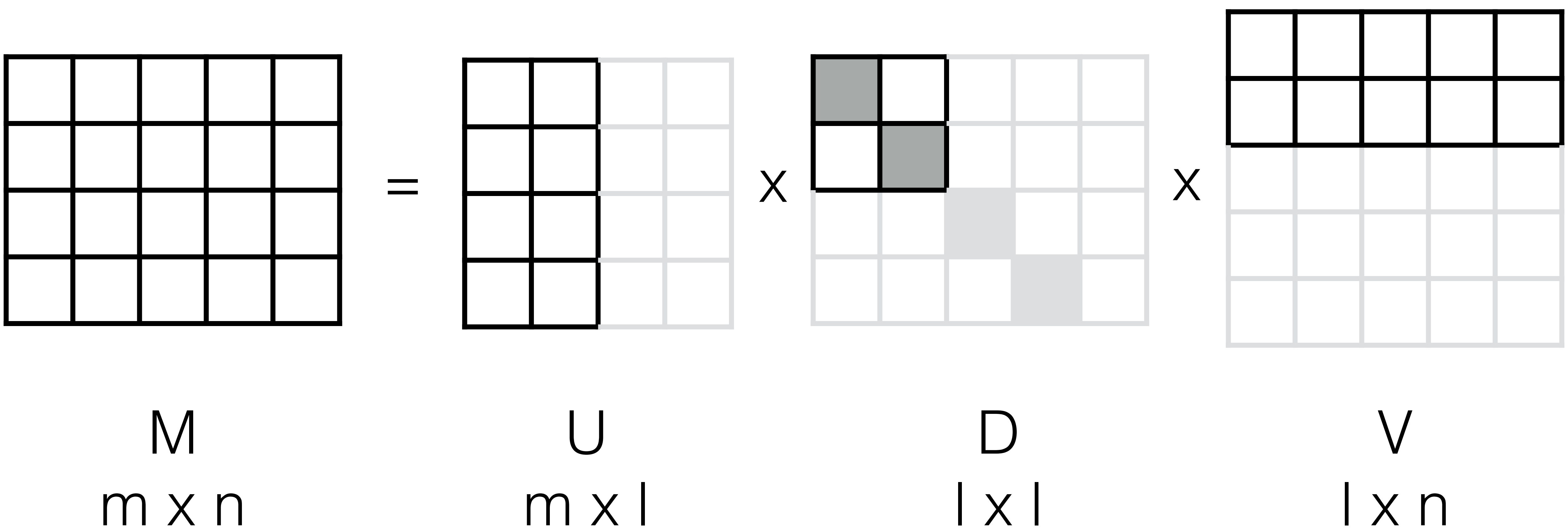
V
l x n

Truncated Singular Value Decomposition (SVD)



keep only first l components

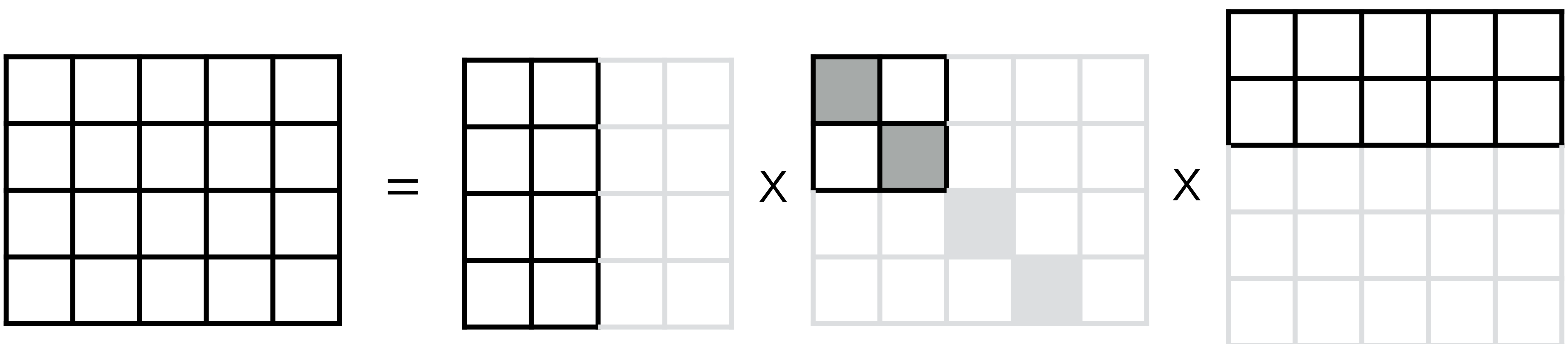
Truncated Singular Value Decomposition (SVD)



keep only first l components

"best l-rank approximation of M"

Truncated Singular Value Decomposition (SVD)



M
m x n

||M − UDV||² as small as possible

keep only first l components

"best l-rank approximation of M"

Dimensionality Reduction

- "Low Rank Assumption": we typically assume that our features contain a large amount of redundant information

- We can throw away a lot of principle components without losing too much of the signal needed for our task

Clicker Question!

Matrices IRL

Matrices IRL

- Data is noisy, so M is most likely full-rank

Matrices IRL

- Data is noisy, so M is most likely full-rank
- We assume that M is *close to* a low rank matrix, and we approximate the matrix it is close to

Matrices IRL

- Data is noisy, so M is most likely full-rank
- We assume that M is *close to* a low rank matrix, and we approximate the matrix it is close to
- Viewed as a “de-noised” version of M

Matrices IRL

- Data is noisy, so M is most likely full-rank
- We assume that M is *close to* a low rank matrix, and we approximate the matrix it is close to
- Viewed as a "de-noised" version of M
- "Original matrix exhibits redundancy and noise, low-rank reconstruction exploits the former to remove the latter"*

*Matrix and Tensor Factorization Methods for Natural Language Processing. (ACL 2015)

Matrices IRL

Matrices IRL

- Data is also often incomplete…missing values, new observations, etc.

Matrices IRL

- Data is also often incomplete…missing values, new observations, etc.
- Can we use SVD for this?

Matrices IRL

- Data is also often incomplete…missing values, new observations, etc.
- Can we use SVD for this?
- Yes! Though we need to make a few changes…

Matrix Completion

	roma	ballad of buster scruggs…	mud-bound	to all the boys i loved before	okja
user1	1	0	1		
user2	0			0	1
user3	1	0		1	0
user4		1	0		
user5					1

Matrix Completion

	roma	ballad of buster scruggs…	mud-bound	to all the boys i loved before	okja
user1	1	0	1	1	0
user2	0			0	1
user3	1	0	1	1	0
user4		1	0		
user5					1

"people also liked…"

NETFLIX

Netflix Prize

COMPLETED

Home | Rules | Leaderboard | Update

Leaderboard

Showing Test Score. Click here to show quiz score

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries !	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07

Matrix Completion

$$M \approx UDV = M'$$

Matrix Completion

$$\boxed{M} \approx UDV = \boxed{M'}$$

original

completed

Matrix Completion

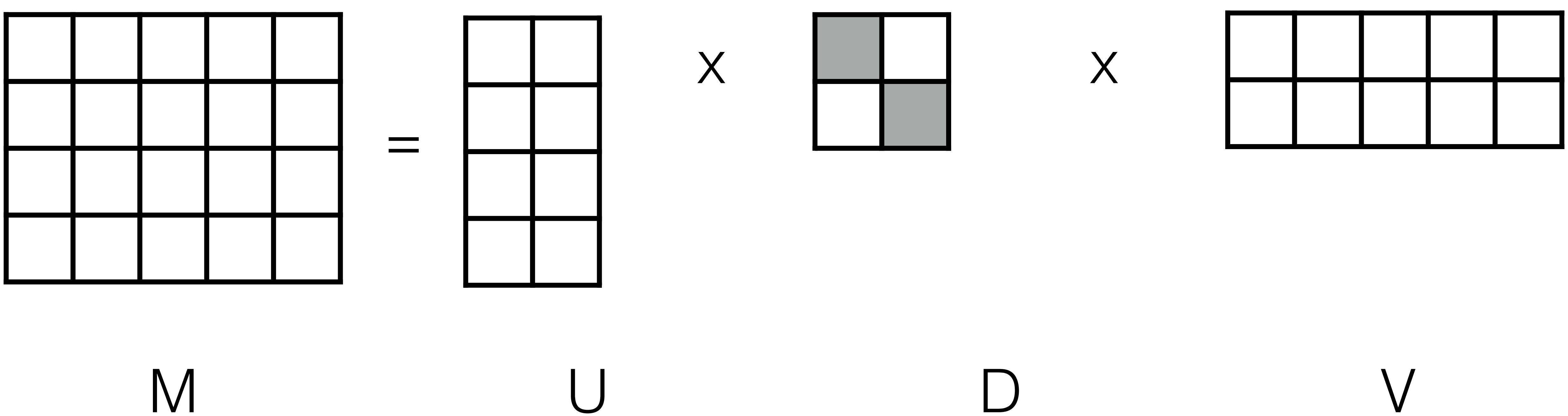
$$\boxed{M} \approx UDV = \boxed{M'}$$

original

completed

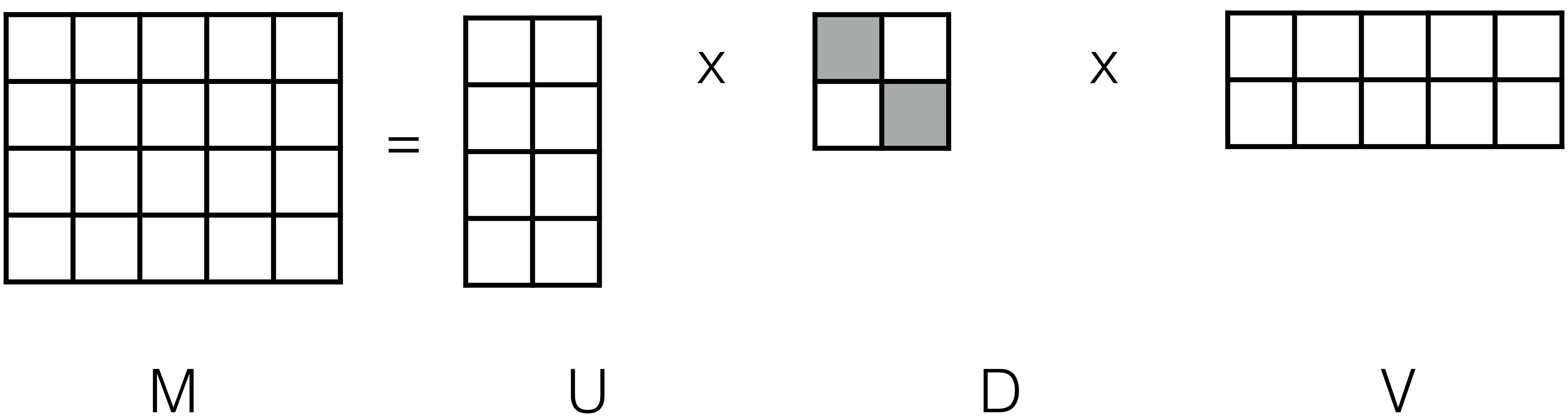
problems?

Matrix Completion



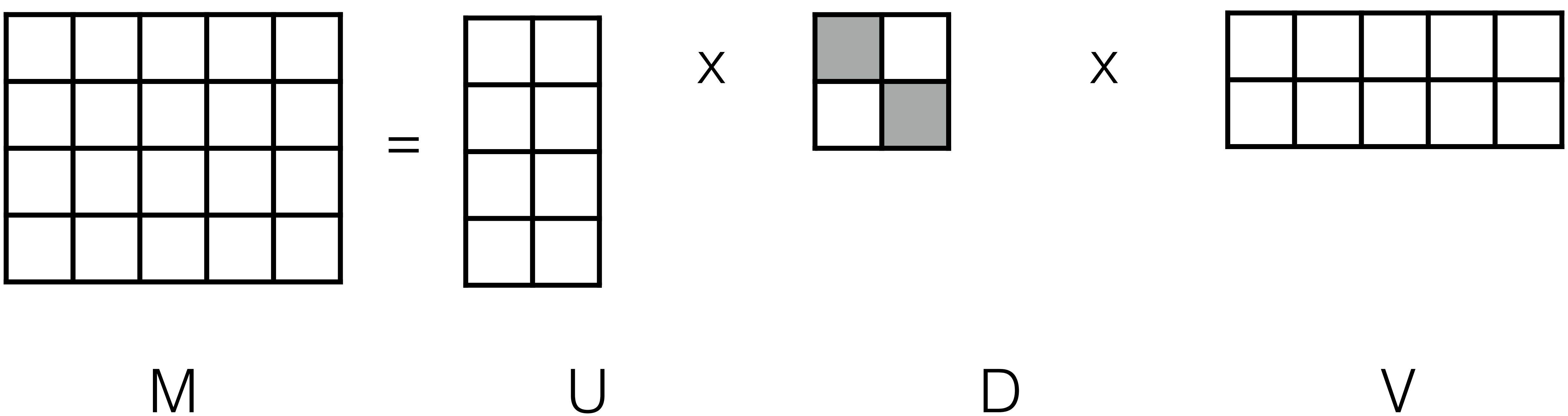
Exact SVD assumes M is complete...

Matrix Completion

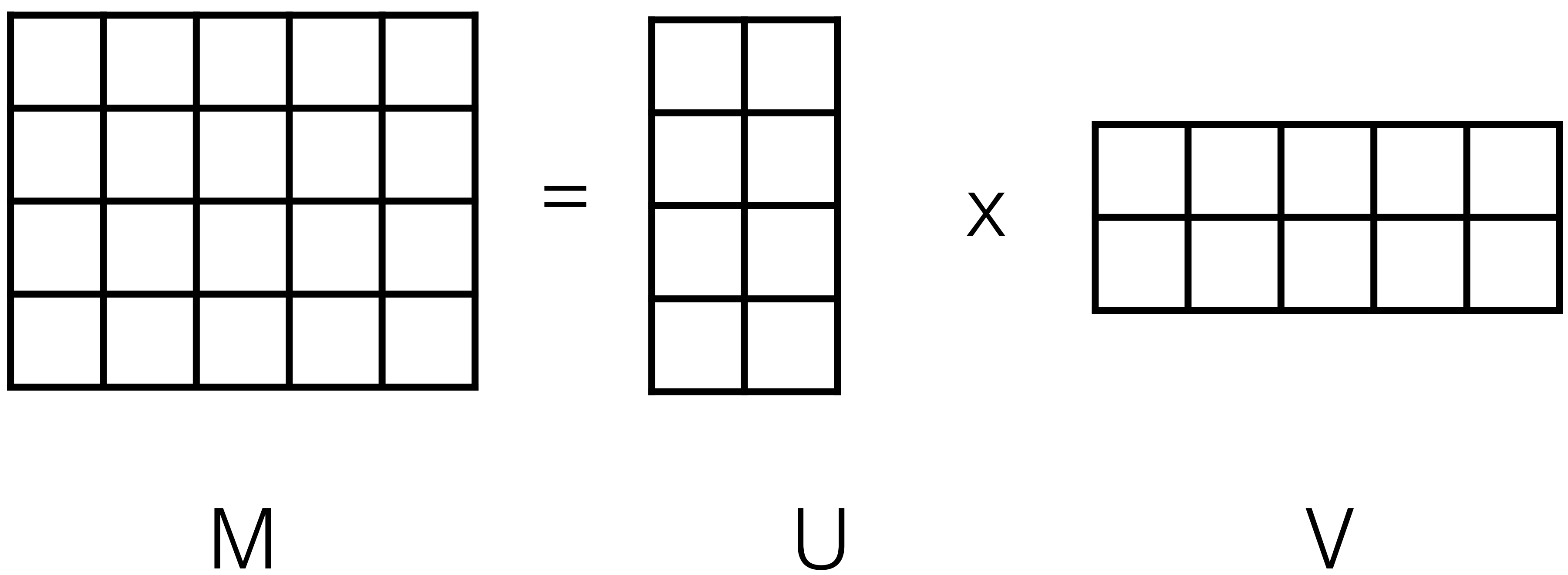


...just gradient descent that MF!

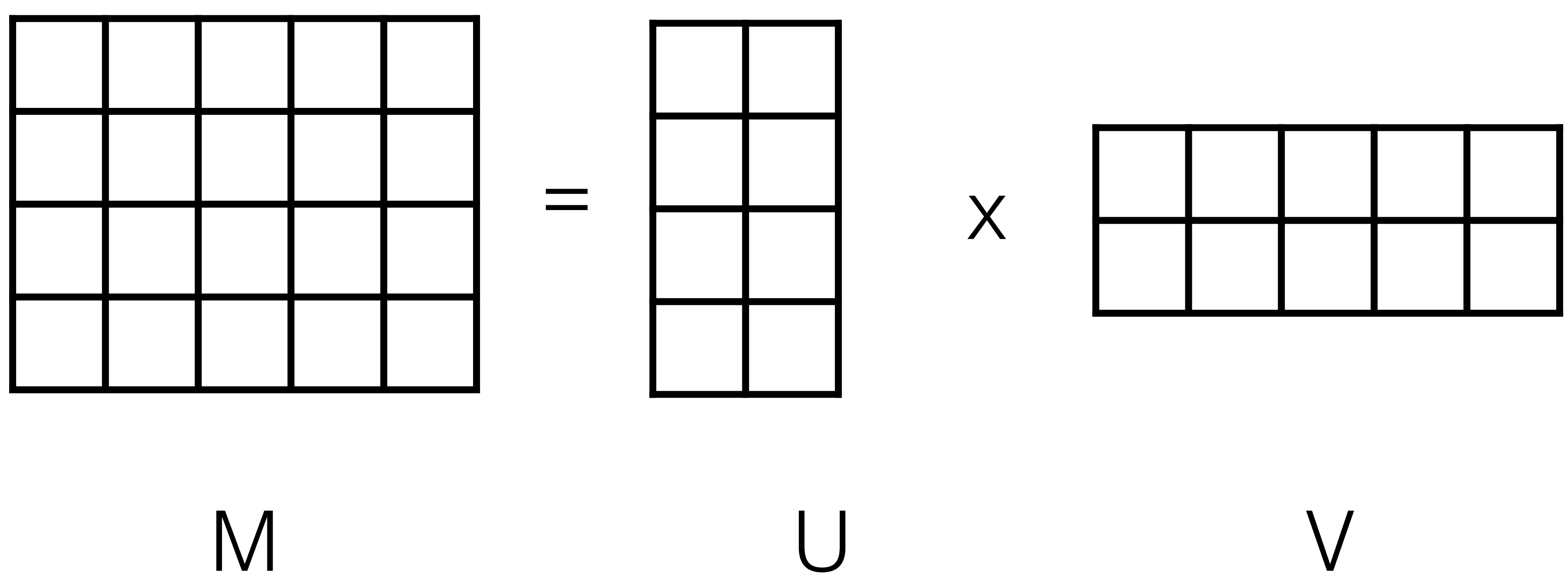
MF with Gradient Descent



MF with Gradient Descent

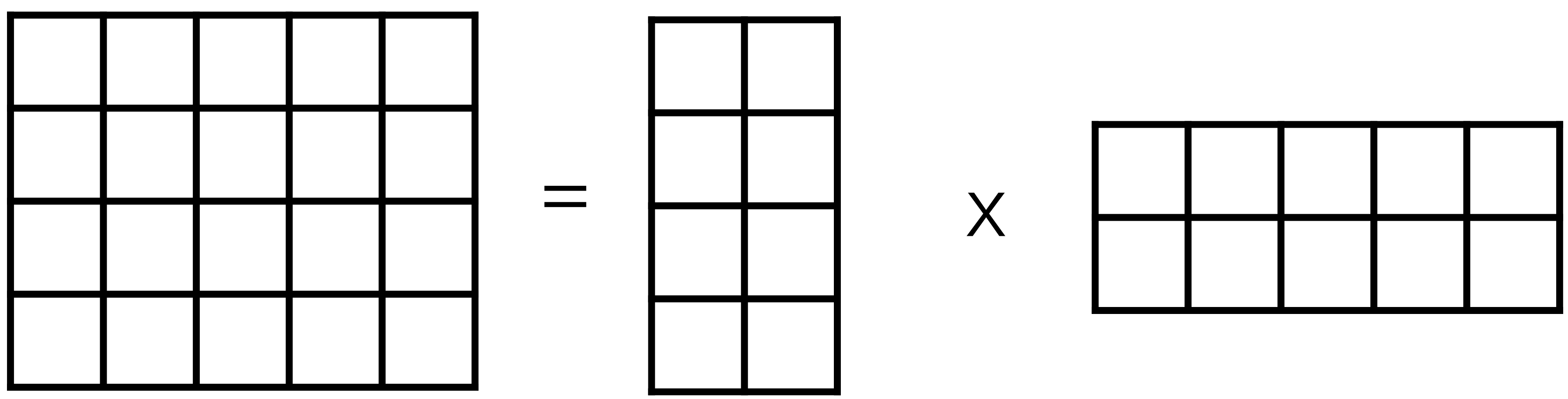


MF with Gradient Descent



Not properly SVD
(fewer guarantees, e.g. components not orthonormal) but good enough

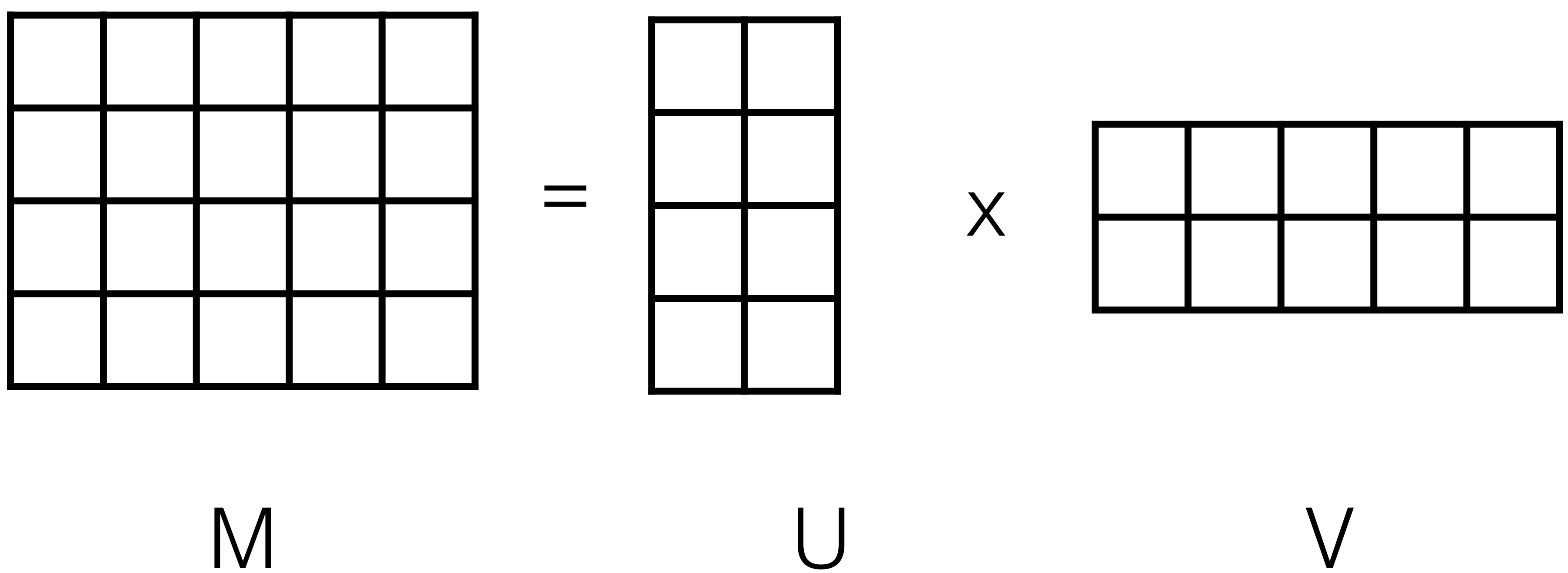
MF with Gradient Descent



M U V

$$\min_{U,V} \sum_{ij} (M_{ij} - u_i \cdot v_j)^2$$

MF with Gradient Descent



$$\min_{U,V} \sum_{ij} (M_{ij} - u_i \cdot v_j)^2$$

But! Only consider cases when M_{ij} is observed!

Clicker Question!

Topic Models

Can you elaborate on exactly what the directions are in part 2 step 3, the stencil code does not quite imply what we are supposed to do…

When I try to display dots from part 2 on my mac (tried chrome, firefox, and safari), the elements do not appear in the html.

Changes I make to the nations.js file do not affect any of the html in after I load the nations.html file

Topic Models

Can you elaborate on exactly what the directions are in part 2 step 3, the stencil code does not quite imply what we are supposed to do…

When I try to display dots from part 2 on my mac (tried chrome, firefox, and safari), the elements do not appear in the html.

Changes I make to the nations.js file do not affect any of the html in after I load the nations.html file

instructions: stencil, instructions, part, step, rubric, handin…

UI: html, javascript, debug, display, elements…

systems: mac, windows, linux, chrome, firefox, os…

fillers: I, you, when, the, and, a

Topic Models

"Latent Semantic Analysis" (LSA)

$$P(w_i) = \sum_{j=1}^{T} P(w_i \mid z_i = j) P(z_i = j)$$

Topic Models

"Latent Semantic Analysis" (LSA)

$$P(w_i) = \sum_{j=1}^{T} P(w_i \mid z_i = j) P(z_i = j)$$

words are determined by topic
(and are conditionally independent of each other)

Topic Models

“Latent Semantic Analysis” (LSA)

$$P(w_i) = \sum_{j=1}^{T} P(w_i \mid z_i = j) P(z_i = j)$$

documents are a distribution over topics c Models

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

U

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

D

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

V

c Models

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

component = "topic"

U

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

D

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

V

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

c Models

component = "topic" = distribution over words

U

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

D

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

V

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

c Models

	the	cong ress	parli ame	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

document = distribution over topics

d1	-0.60	-0.39	0.70	0.00
d2	-0.48	0.50	-0.12	-0.71
d3	-0.43	-0.58	-0.69	0.00
d4	-0.48	0.50	-0.12	0.71

U

3.06	0.00	0.00	0.00	0.00
0.00	1.81	0.00	0.00	0.00
0.00	0.00	0.57	0.00	0.00
0.00	0.00	0.00	0.00	0.00

D

the	cong ress	parlia ment	US	UK
-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

V

Topic Models

Factorization of the term-document matrix

	the	congress	parliament	US	UK
doc1	1	1	1	1	0
doc2	1	0	1	0	1
doc3	1	1	0	1	0
doc4	1	0	1	0	1

More on Thursday!

Word Embeddings

Factorization of the term-context matrix

	the	congress	parliament	US	UK
the	1	1	1	1	1
congress	1	1	0	1	0
parlaiment	1	0	1	1	1
US	1	1	1	1	0
UK	1	0	1	0	1

More on Thursday!

Word Embeddings

	the	con-gress	parlia-ment	US	UK
the	1	1	1	1	1
congress	1	1	0	1	0
parlaiment	1	0	1	1	1
US	1	1	1	1	0
UK	1	0	1	0	1

Embeddings!

=

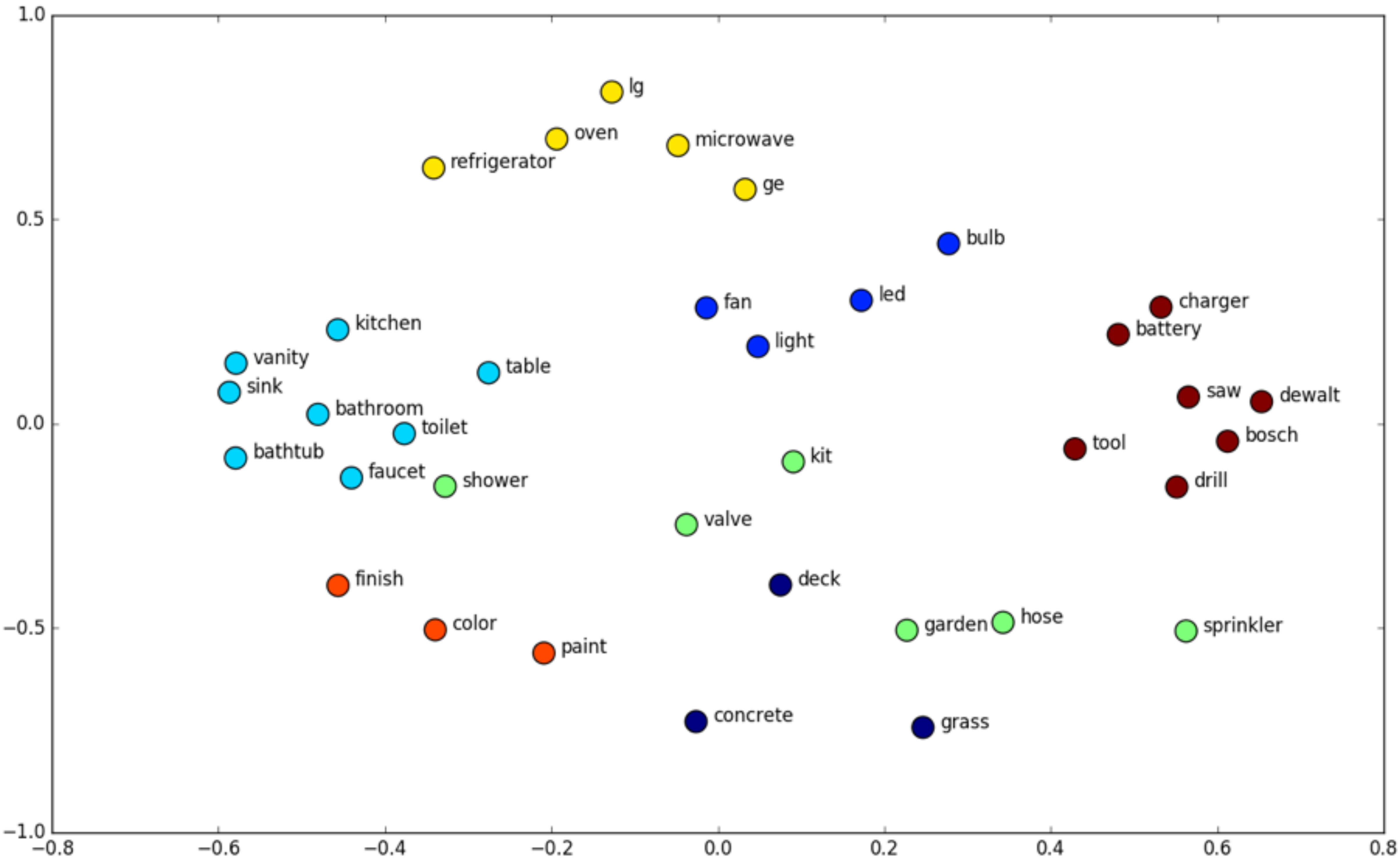
the	-0.60	-0.39	0.70	0.00
congress	-0.48	0.50	-0.12	-0.71
parliament	-0.43	-0.58	-0.69	0.00
US	-0.48	0.50	-0.12	0.71
UK	0.02	0.79	0.02	-0.44

×

-0.65	-0.34	-0.51	-0.34	-0.31
0.02	-0.54	0.34	-0.54	0.56
-0.42	0.02	0.79	0.02	-0.44
-0.63	0.27	0.00	0.37	0.63
-0.04	0.73	0.00	-0.68	0.04

More on Thursday!

Word Embeddings



Useful Resources/ References

- https://github.com/uclmr/acl2015tutorial/
- https://web.stanford.edu/~jurafsky/li15/lec3.vector.pdf
- https://arxiv.org/pdf/1404.1100.pdf
- https://towardsdatascience.com/pca-and-svd-explained-with-numpy-5d13b0d2a4d8
- http://nicolas-hug.com/blog/matrix_facto_3
- https://machinelearningmastery.com/singular-value-decomposition-for-machine-learning/
- http://cocosci.princeton.edu/tom/papers/SteyversGriffiths.pdf