Day 19: Recommender Systems and Dimensionality Reduction

Recommender Systems

Goal: recommend things

products, movies, music, websites, etc.

2 basic types:

1. content-based
2. collaborative filtering

Content-Based Systems

- Recommend based on explicit features
- Users and items share the same feature space

Example (unsupervised):

Movies: ("items")

	Comedy	<120 mins	Korean?	Take place in Canada
1	1	0	1	0
2	0	1	1	1
3	1	1	1	1

Users:

A	1	0	1	0
B	0	1	1	1

Dot product between a user and a movie gives similarity.

$$\text{sim}(\text{movie } 1, \text{ user } A) = 1\cdot 1 + 0\cdot 0 + 1\cdot 1 + 0\cdot 0 = 2$$

$$\text{sim}(\text{movie } 1, \text{ user } B) = 1\cdot 0 + 0\cdot 1 + 1\cdot 1 + 0\cdot 0 = 1$$

Example (supervised)

Movies: same as above

Users:

A	θ_{1A}	θ_{2A}	θ_{3A}	θ_{4A}
B	θ_{1B}	θ_{2B}	θ_{3B}	θ_{4B}

Train a model to either:

- predict the user's rating (regression)
- predict whether user will like movie (class...)

Collaborative Filtering

Basic idea: predict rating based on either similar items or similar users

Two Variations:

1. user-user
2. item-item

User-User

1. find users w/ similar rating patterns

(use KNN, K-means, etc.)

2. use ratings of similar users to calculate rating for current users

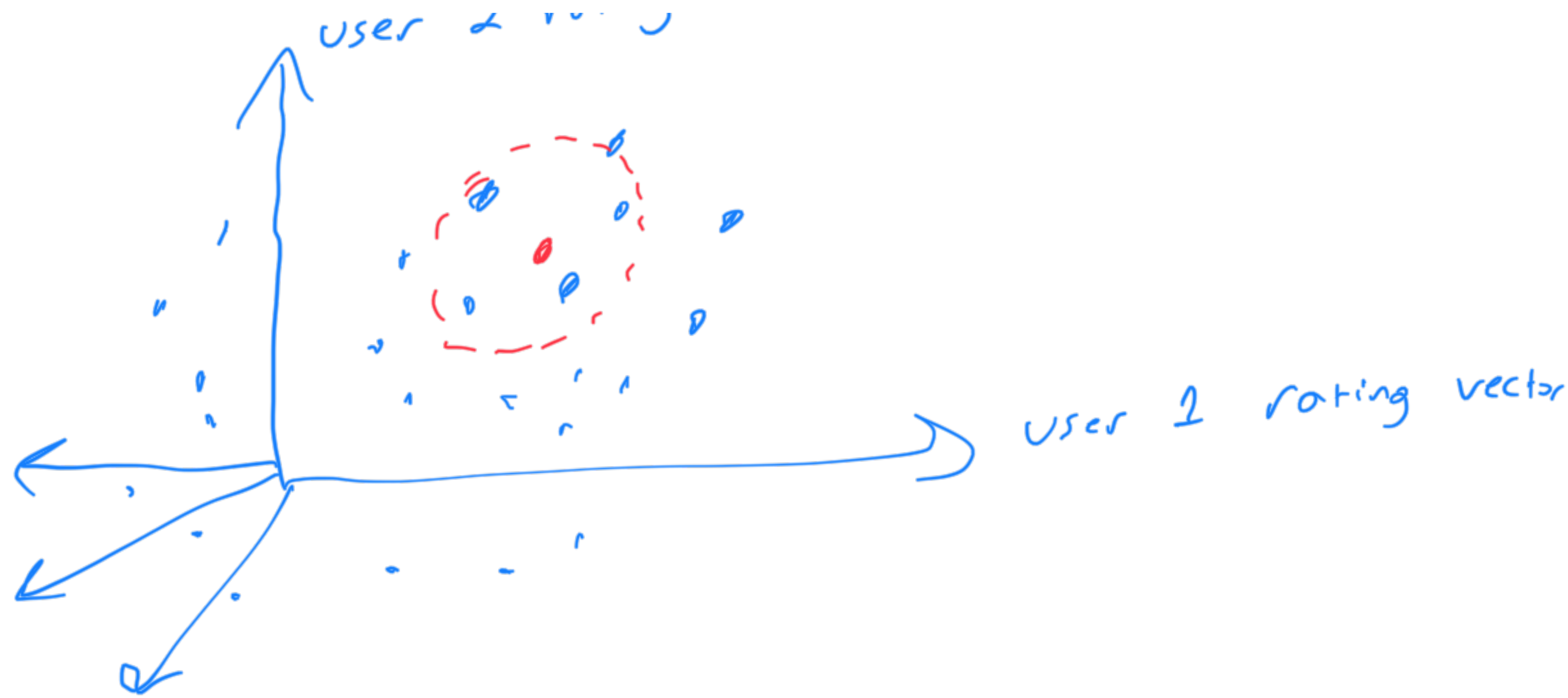
3. suggest items according to highest predicted ratings

Item - Item

"users who like X also like Y"

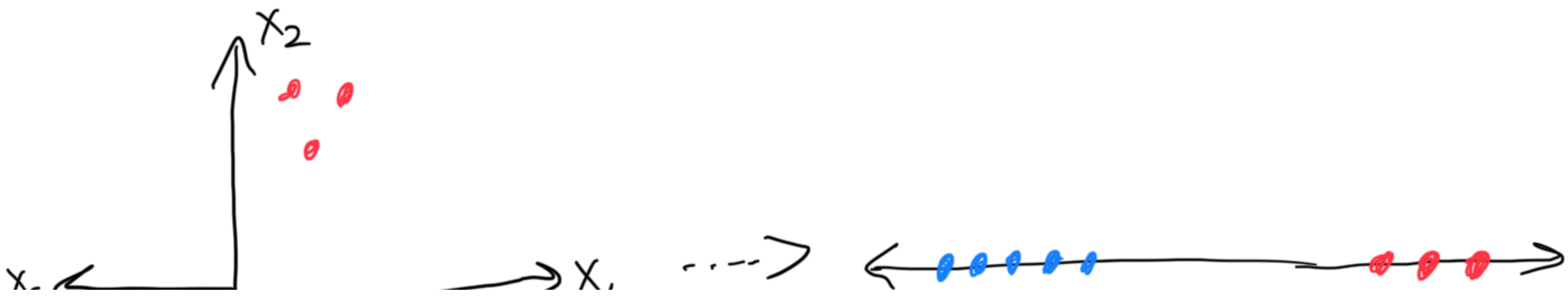
Find items similar to an item the user already bought based only on the ratings for that item

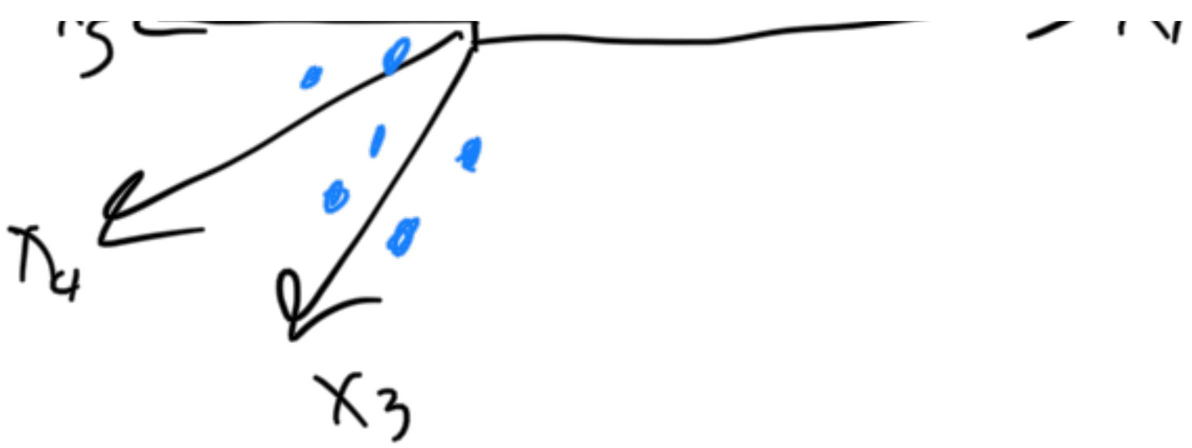
rating vector



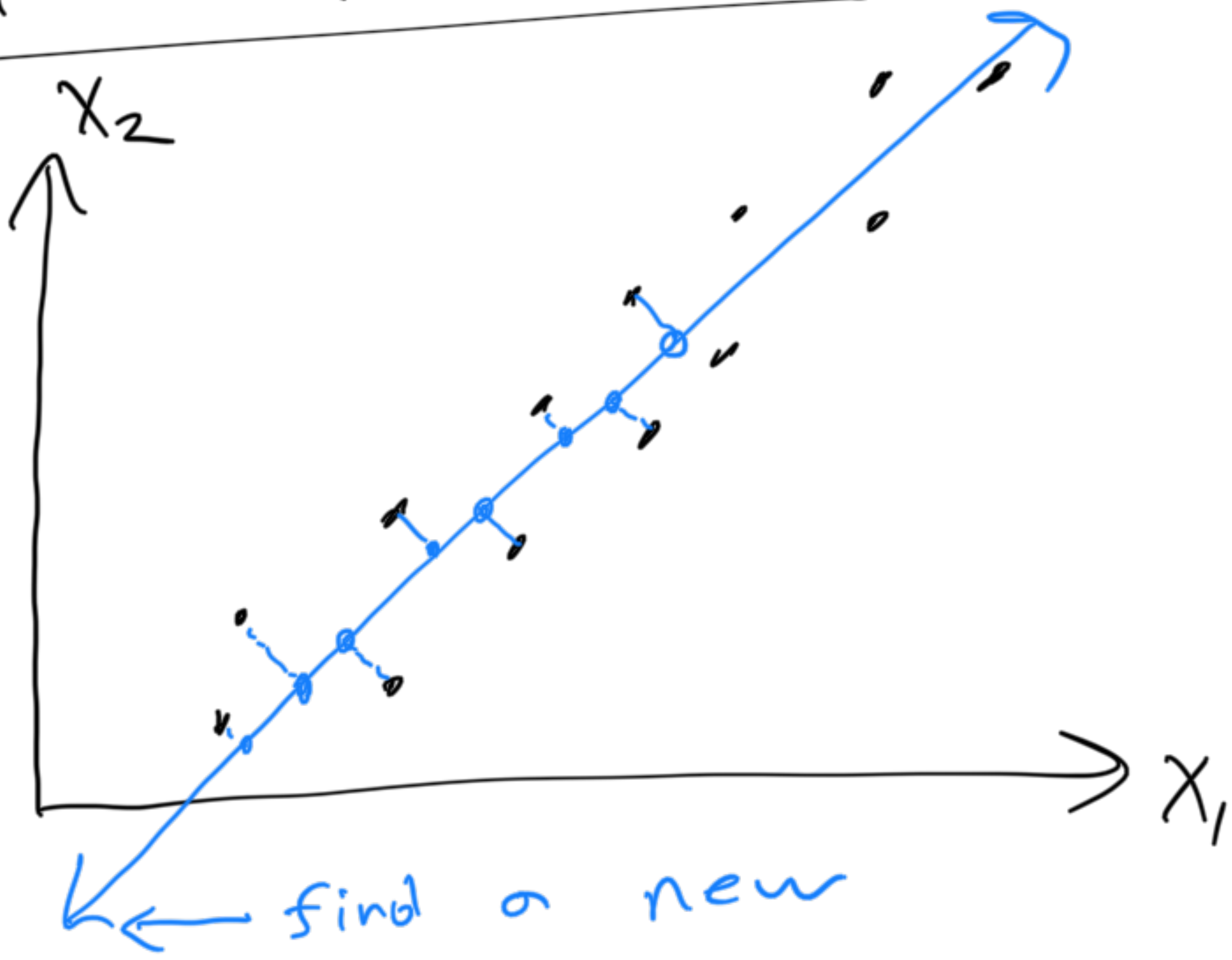
These are all general ideas; there are many variations.

Dimensionality Reduction

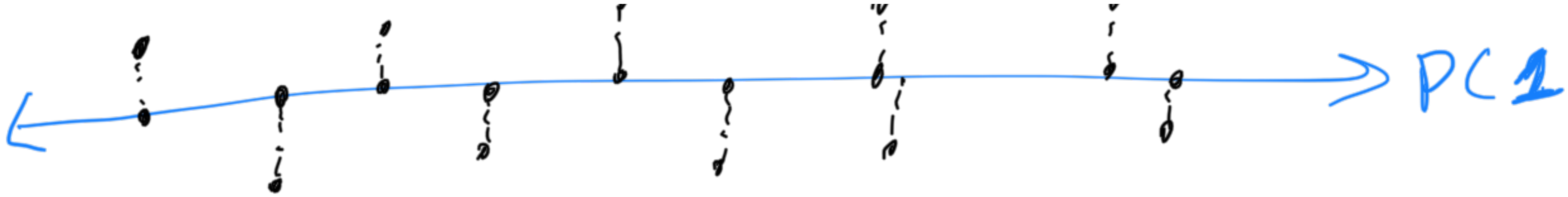




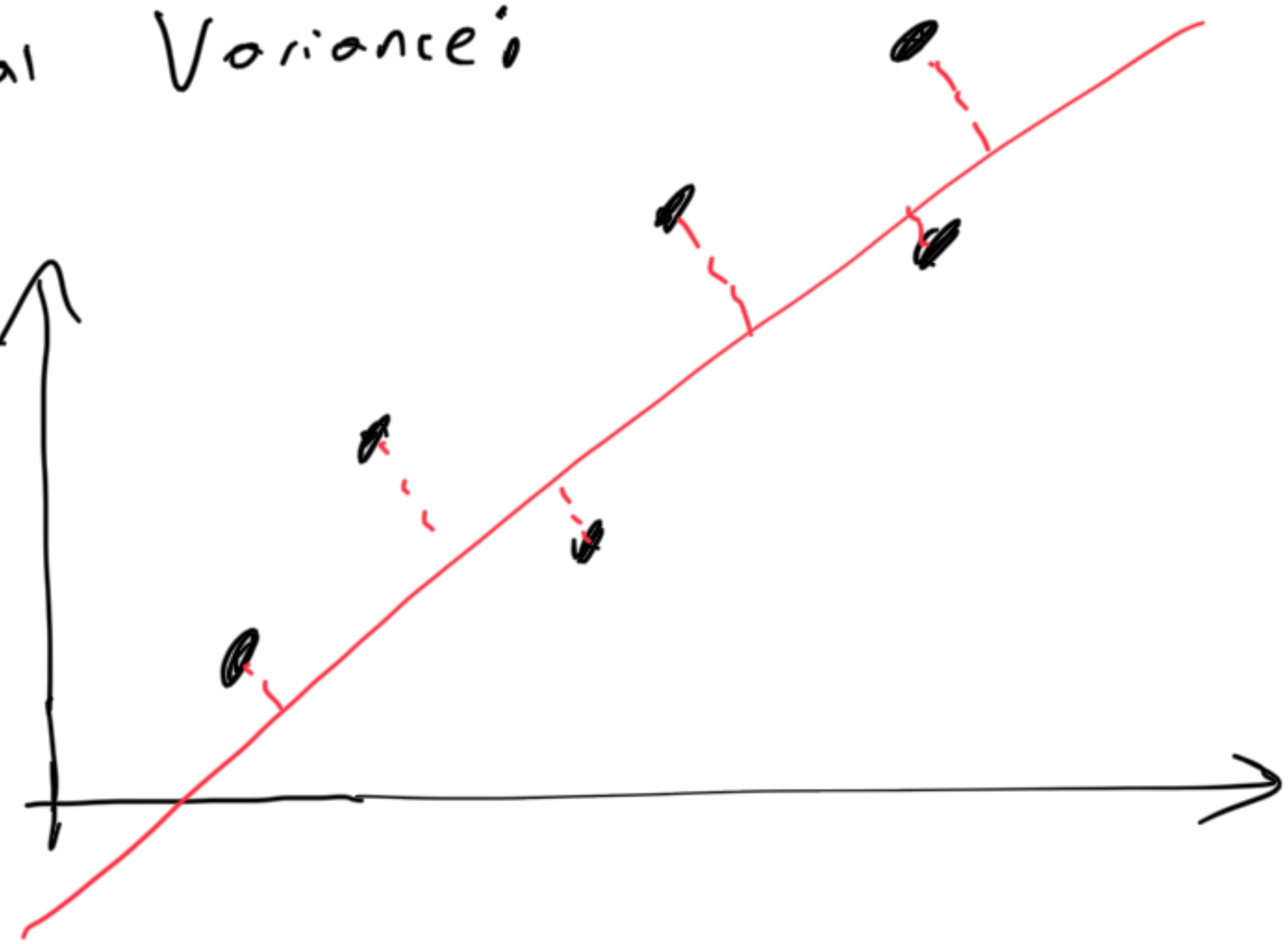
Principal Component Analysis (PCA)



find a new axis of the original data which describes the maximal variance



Maximal Variance:

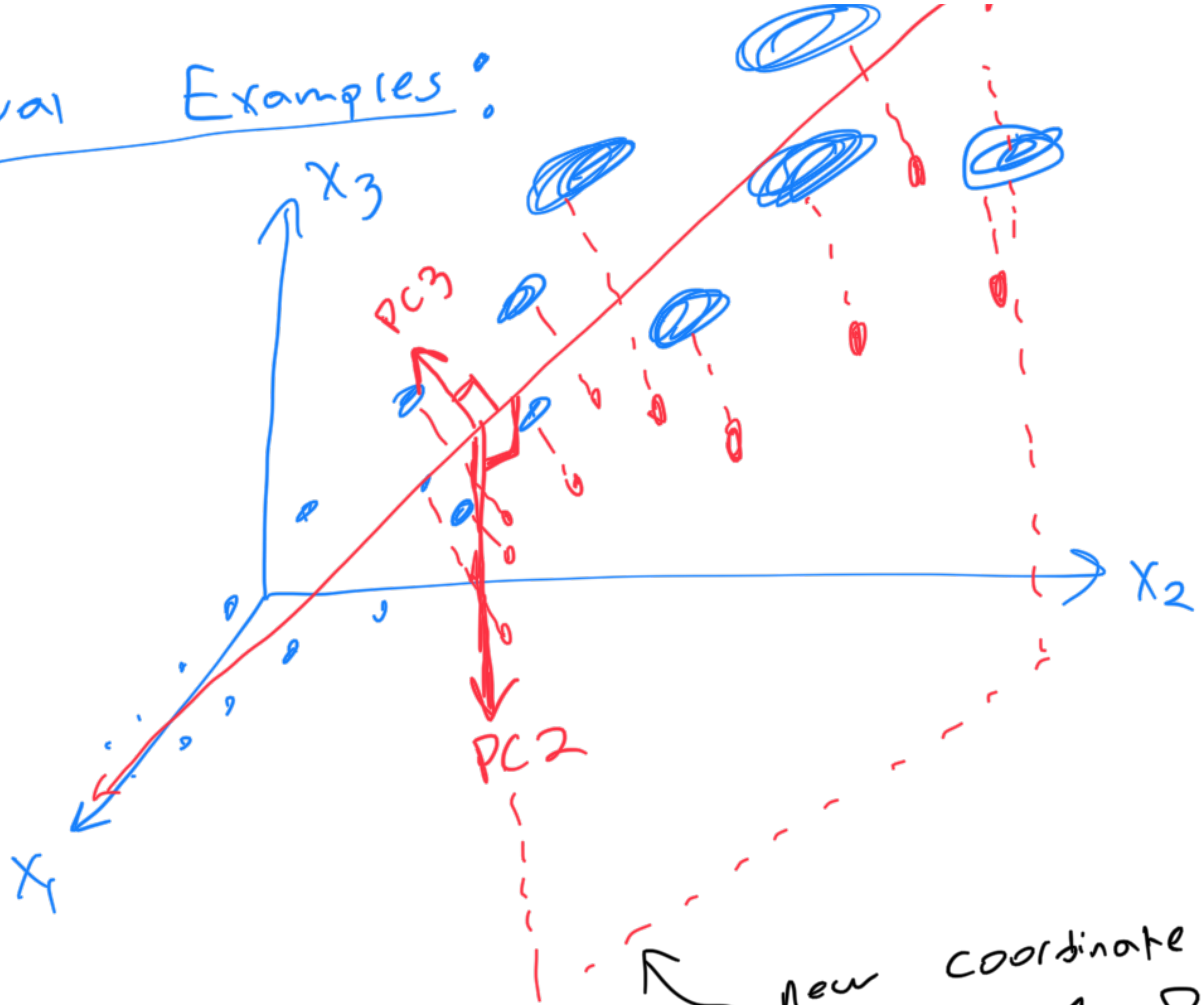


The new axis is called "principal component 1" or PC1

PC2 is an ~~orthogonal~~ axis to PC1 which has the second highest variance

(bigger points

Visual Examples:

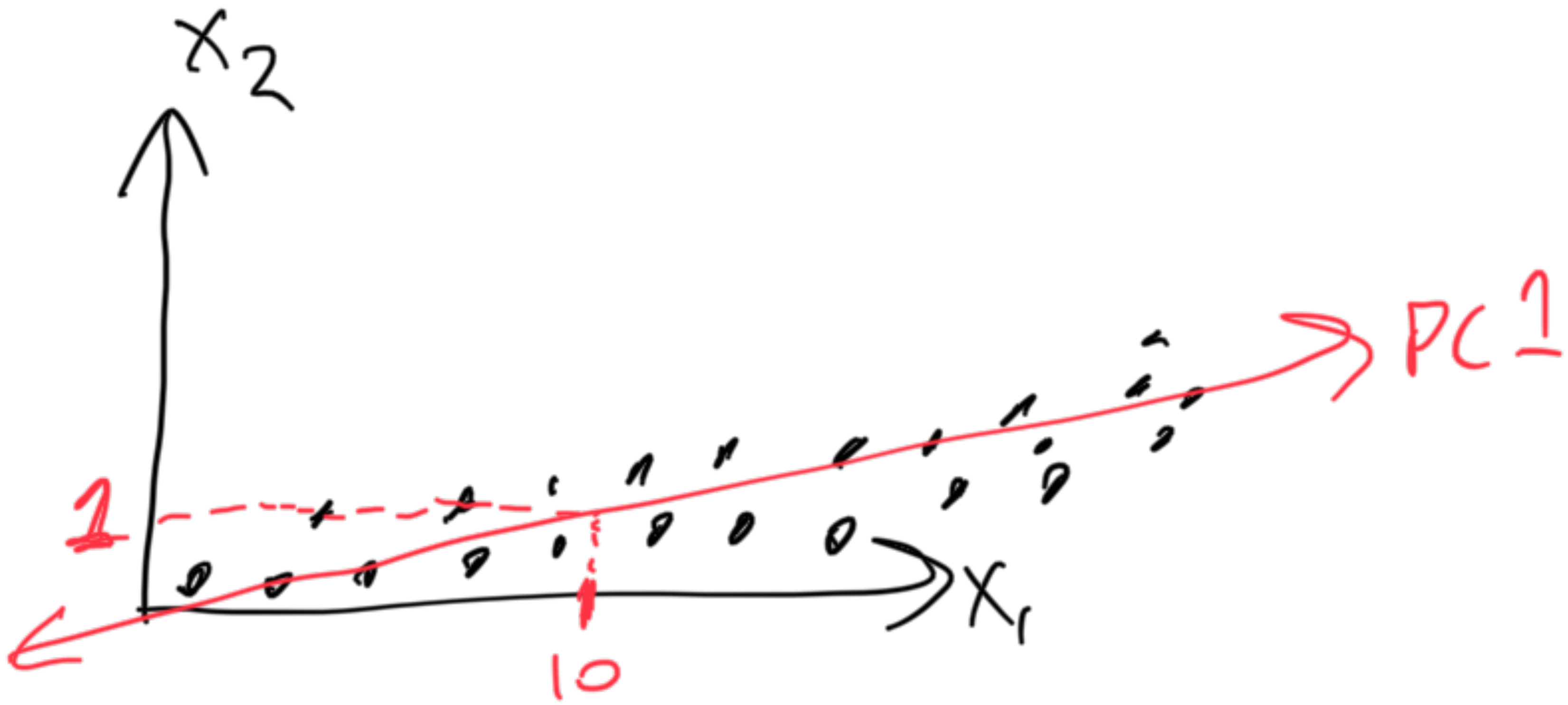


depict closer points to the observer of these notes in 3D space)

new coordinate system with axes PC1, PC2, with each point in the original space projected onto this coordinate space, is the reduced dimensional data

…ding to PCA

according to



PC1, in this case, is 10 parts X_1 and 1 part X_2

In other words, X_1 is 10x more important/variable than X_2

How to calculate PCs

(1) standardize data to make each feature unit-less:

$$\text{New Value} = \frac{\text{Old Value} - \text{Mean Value of that feature}}{\ldots}$$

New var N ← # data points

(population)
② Calculate covariance matrix

Covariance = how 2 variables vary with respect to each other

$$\text{Covariance}(A, B) = \text{Cov}(A, B) = \frac{\sum (a_i - \bar{a})(b_i - \bar{b})}{N}$$

Covariance matrix for 3 features a, b, c =

$$\begin{bmatrix} \text{cov}(a,a) & \text{cov}(a,b) & \text{cov}(a,c) \\ \text{cov}(b,a) & \text{cov}(b,b) & \text{cov}(b,c) \\ \text{cov}(c,a) & \text{cov}(c,b) & \text{cov}(c,c) \end{bmatrix}$$

Note: you can calculate covariance matrix simply as: $X^T X$

(if dataset is standardized)

the dataset

Why?

$$X = \begin{array}{c} \\ dp\ 1 \\ dp\ 2 \\ dp\ 3 \end{array} \begin{array}{cc} \text{feat } 1 & \text{feat } 2 \\ \hline \end{array}$$

$$X = \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$$

$$X^T X = \begin{bmatrix} a & c & e \\ b & d & f \end{bmatrix} \begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix}$$

$$= \begin{bmatrix} a \cdot a + c \cdot c + e \cdot e = cov(\text{feat } 1, \text{feat } 1) & \dots \\ = cov(\text{feat } 1, \text{feat } 2) & \\ \dots & \dots \end{bmatrix}$$

(3) Calculate the eigenvalues/eigenvectors of the covariance matrix

$$Av = \lambda v$$

original matrix

eigenvector

eigenvalue

The eigenvectors v are the PCs, with the corresponding λ describing the ranking of the PC

Why does calculating the eigenvalues/vectors of the cov matrix give the direction of maximal variance:

Projecting a point x_i onto vector u:

$$\frac{u^T x_i}{\cancel{\|u\|}}$$

To make things easier, we will make u a unit vector

Mean of projections:

$$u^T \bar{x}$$

Variance of projections:

$$\frac{1}{n}\sum_{i=1}^{n}\left(u^T x_i - u^T \bar{X}\right) = \ldots$$

$$= \frac{1}{n}\sum_{i=1}^{n} u^T (X_i - \bar{X})(X_i - \bar{X}) u$$

$$= u^T \underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}_{\text{covariance matrix} = S} u$$

$$= \underbrace{u^T S u}_{\text{variance of the projections}}$$

covariance matrix

So: Solve $\underbrace{\max u^T S u}_{\text{maximize the}}$ such that $u^T u = 1$

Variance of
the projections

Solve using Lagrange multiple:

$$\frac{d}{du} u^T S u + \lambda(1 - u^T u) = 0$$

$$\underbrace{Su = \lambda u}_{\text{eigenvector/value definition}}$$

$$\underbrace{\lambda = u^T S u}_{\text{biggest eigenvalue gives us the biggest variance of the projected data}}$$

How to project data onto PC space?

$$U^T X$$

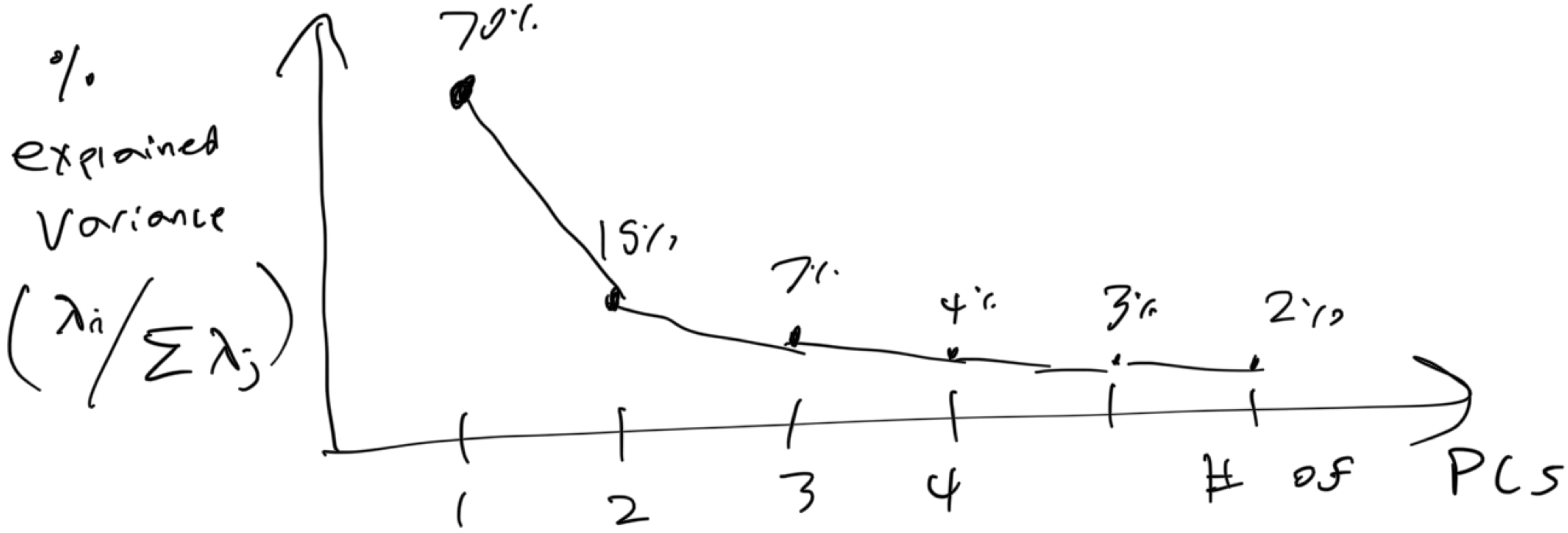
PC

original data

How many PCs to use?

explained variance of a PC $= \dfrac{\lambda_i}{\sum_j \lambda_j}$

Scree Plot:


% explained variance
$(\lambda_i / \sum \lambda_j)$
70%
15%
7%
4%
3%
2%
1
2
3
4
# of PCs