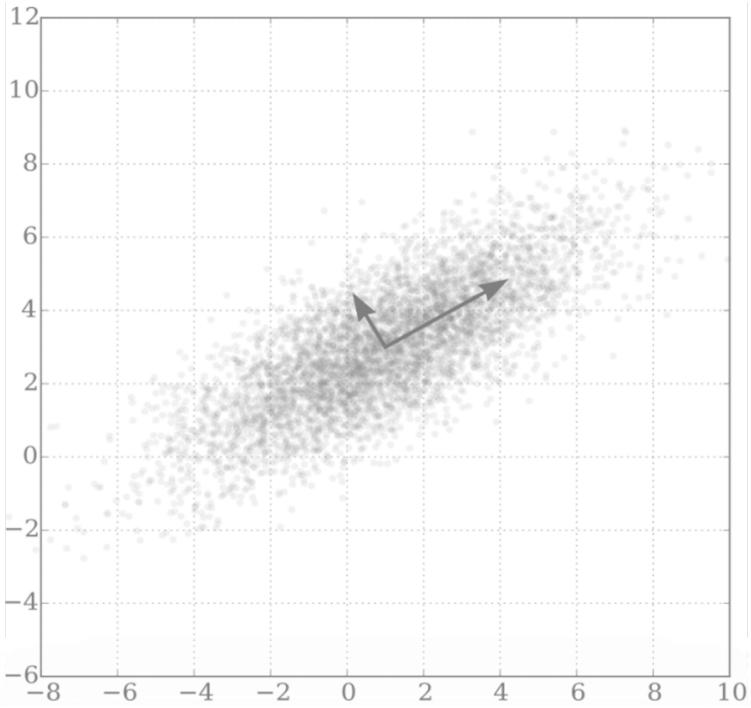
Lecture 15:

Principal Component Analysis



CS 111: Intro to Computational Science

Spring 2023

Ziad Matni, Ph.D.

Dept. of Computer Science, UCSB

Administrative

- ***Reminder: Next week Monday (5/29) is a University Holiday***

- Current homework due today
- New homework out later today

- Quiz 6 on Wednesday: Lectures 13 and 14 (SVD part 2, Covariance)

- Today's lecture is mostly demonstration and will be recorded

Variability Explanation of PCs

- Since we know the principal components (PCs), it would be helpful to know *how much **variability** in the data that each PC explains*

- Generally, whenever you ***model*** data, you'd like that model to "explain" as much of the variability in the data as it can

- If you're going for a reduction down to 2 dimensions (useful), you'd like your top 2 PCs to explain a majority of the variability

Variability Explanation of PCs

Thanks to SVD, this is easy to figure out!

- Technique: Find out how much each singular value is weighed against the total sum of all singular values
  - Recall: singular values, in this context, are the same as the eigenvalues

- Let's say there are ***n*** singular values, therefore….
- **Variation that PC_i accounts for** = **Variation that eigenvalue *i* accounts for**

$$= \frac{\sigma_i}{\sum_{k=0}^{n} \sigma_k}$$

*Best visualized with a **scree plot** (will show in the demo)*

So What Then?

Let’s run through an exercise that allow us to use PCA to visualize a data set

Quick! To the Python-mobile!

Your TO DOs!

- Assignment 07 due tonight
- Quiz 6 on Wednesday
  - Lectures 13 and 14 (SVD part 2, Covariance)

</LECTURE>