EECS 16A
Voice Recognition 1

Welcome! We'll be starting at Berkeley Time.

Semester Outline

Shazam

Imaging

Acoustic Positioning

Voice Recognition

Today's Agenda

- Voice Classification Problem Overview
- SVD / PCA Conceptual Review
- Using SVD / PCA for Voice Classification
- Tips for Best Performance

Voice Recognition Problem

- We have a set of:
  - Words
  - 40 recordings of someone speaking each word
- We want to come up with a model that can be given a new recording of the spoken word and predict which word it is!

- In other words, given a training dataset, “train” a classifier model that takes audio signals.

Organization of our Data

- Each recording is a vector of some numbers that represent our sampled audio signal
- Given our data, what can we do to classify a new recording? What are some features that we can compare?

```
Recording 0 =  [-64, 30, 24, 52, …]
Recording 1 =  [-54, 86, 78, 64, …]
Recording 2 =  [32, 64, 50, -110 …]
Recording 3 =  [-40, 70, 4, -36, …]
⋮
```

Features

- In this lab, we are going to stack all the recordings into a matrix where the row is each recording and each audio sample (column) is a feature (things that distinguish one datapoint from another).

feature		0	1	2	3	
Recording 0 =	[	-64,	30,	24,	52,	…]
Recording 1 =	[	-54,	86,	78,	64,	…]
Recording 2 =	[	32,	64,	50,	-110	…]
Recording 3 =	[	-40,	70,	4,	-36,	…]

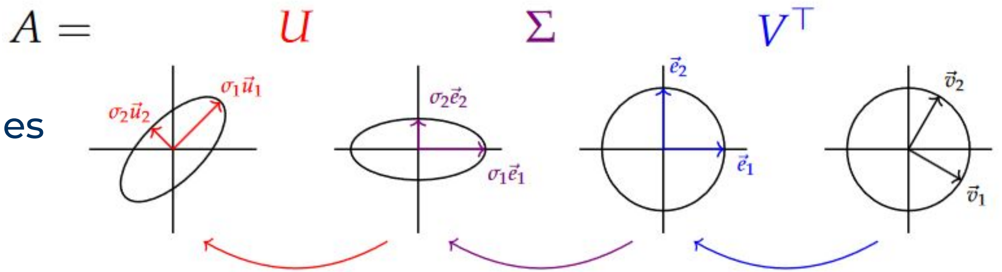
⋮

Singular Value Decomposition (SVD)

- What is SVD?
  - **Decompose a matrix** into 3 separate matrices with unique properties
- U
  - Orthonormal eigenvectors of AA^T
  - A rotation matrix!
- S
  - Diagonal matrix of singular values
  - Think of as a scaling matrix!
- V
  - Orthonormal eigenvectors of A^TA
  - Another rotation matrix!

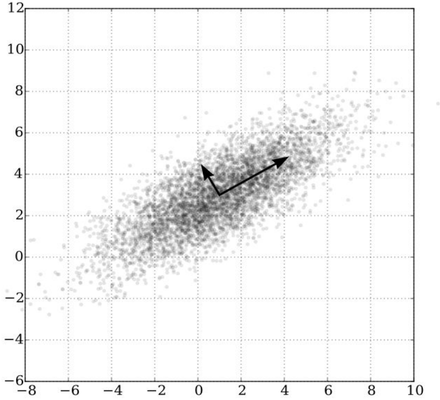
$$A_{m\times n} = U_{m\times m} S_{m\times n} V^T_{n\times n}$$

$$\underset{m\times n}{\overset{\mathrm{A}}{\begin{pmatrix} x_{11} & x_{12} & & x_{1n} \\ & & \ddots & \\ x_{m1} & & & x_{mn} \end{pmatrix}}} = \underset{m\times m}{\overset{\mathrm{U}}{\begin{pmatrix} u_{11} & & u_{m1} \\ & \ddots & \\ u_{1m} & & u_{mm} \end{pmatrix}}} \underset{m\times n}{\overset{\mathrm{S}}{\begin{pmatrix} \sigma_1 & & & 0 \\ & \ddots & & \\ & & \sigma_r & \\ & & & \ddots \\ 0 & & & 0 \end{pmatrix}}} \underset{n\times n}{\overset{\mathrm{V^T}}{\begin{pmatrix} v_{11} & & v_{1n} \\ & \ddots & \\ v_{n1} & & v_{nn} \end{pmatrix}}}$$

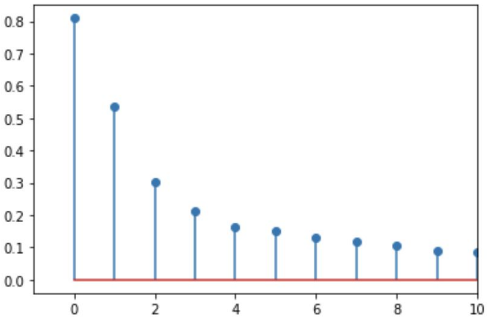


Principal Components Analysis (PCA)

- PCA = Principal Component Analysis
  - **Principal components:** basis vectors that maximize variance in our data
  - Oftentimes, we can capture most of the data's behavior with just a few principal components!
    - **Fewer dimensions is easier to work with**
- How do we compute PCA?
  - Let's use SVD!!!!
  - Take the vectors that correspond with the highest singular values since those are the "most important" transformations of a matrix
  - The principal components of our setup are the vectors from V – **why?**



Principal Components of Data Example



Sigma Values Example

What are our PCA Vectors?

Theorem 9 (Linear Algebra of the Full SVD)
Let $A \in \mathbb{R}^{m \times n}$ have rank $r \leq \min\{m, n\}$. Let $A = U\Sigma V^\top$ be an SVD of A.

(i) $\mathrm{Col}(U_r) = \mathrm{Col}(A)$.

(ii) $\boxed{\mathrm{Col}(V_r) = \mathrm{Col}(A^\top) = \mathrm{Row}(A).}$

(iii) $\mathrm{Col}(U_{m-r}) = \mathrm{Null}(A^\top)$.

(iv) $\mathrm{Col}(V_{n-r}) = \mathrm{Null}(A)$.

(v) $AA^\top = U\Sigma\Sigma^\top U^\top$.

(vi) $A^\top A = V\Sigma^\top \Sigma V^\top$.

(vii) $AV_r = U_r\Sigma_r$.

Each row is one datapoint, we want to capture the variance across all our datapoints!

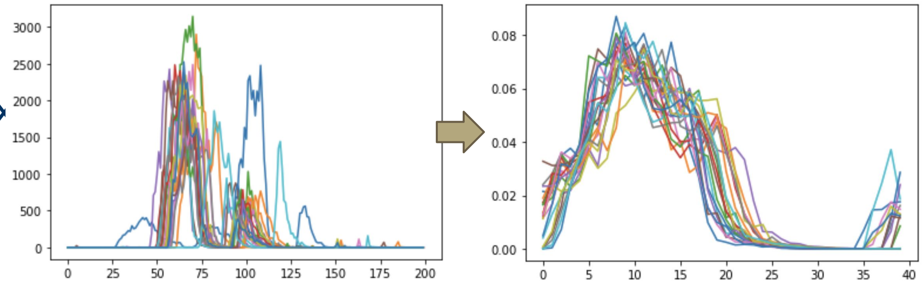
Voice Classification with PCA

Our voice classification system:

1. Data Pre-Processing
2. SVD and PCA Computation
3. Mean Centroid Classification
4. Validation + Hyperparameter Tuning

Data Preprocessing/Word Alignment

- Trim and align each recording to locate and isolate the spoken word
  - ***Threshold*** - percentage of the max value needed to trigger a word start
  - ***Pre-length*** **-** how many timesteps before we hit this threshold did we start speaking the word
  - ***Length*** **-** how long the sample is



Word example: “meat”

Computing PCA via SVD

- Split our data into (70%) training and (30%) testing data
    - Use training data for the steps below, save testing data for Part 6
- Stack the aligned words in a data matrix
- Zero-mean (“demean”) the data using the mean of each timestep (each “feature”) in preparation for SVD

demeaned_A =

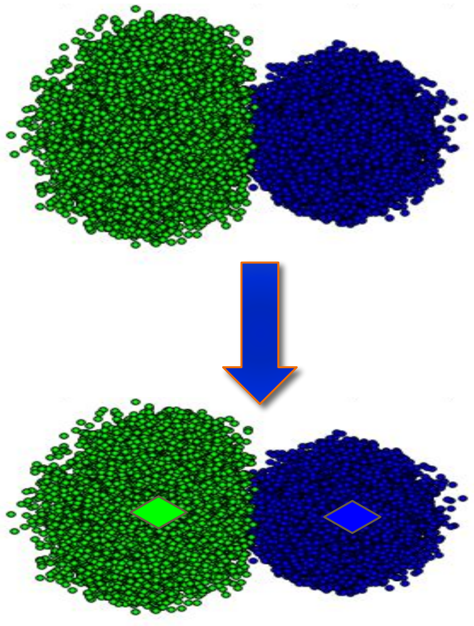
word1_processed_train
word2_processed_train
word3_processed_train
word4_processed_train

\- mean(processed_A)

processed_A

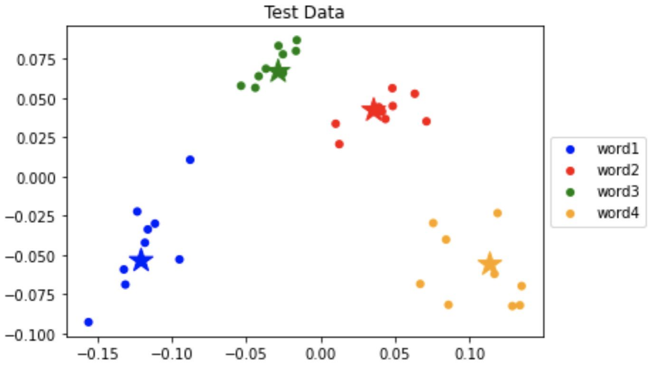
Mean Centroid Classification

- We now have:
  - Labelled training data (70% of Original Data)
  - Axes of most variation (PCA basis vectors)
- Project data onto basis vectors
- Find projected centroid (mean) for each word
  - To classify new points (unlabelled data), project it onto these basis vectors and finding which centroid it is closest to



Verify your Classifier

- Using the 30% of data set aside, verify the accuracy of our classifier
  - Project test data onto our PCA vectors
  - Subtract projected mean vector
  - Assign to closest centroid in 3D space
  - Check if classified centroid is the same as the data label
- Aim for at least 80% accuracy on all words with the test data!
  - In machine learning, it is standard practice to follow this “set aside” method to verify that the classification works and isn’t overfitting



Word Choice Guidelines

- Speech pattern recognition, not word recognition
  - Actual word doesn’t matter, but the speech pattern for the word does!
  - Enunciate syllables well to make clear distinctions
- Generally, try to use words with:
  - Different syllables (e.g. pear, apple, banana, watermelon)
  - Different endings (hard vs. soft, , e.g back vs. shoe)
- Remember the way you say each word, as you’ll have to replicate it later!
  - Can record yourself with your phone so you have a record of how you said each word

Data Collection Guidelines

- We will provide 4 word datasets
- You will collect 40 recordings of one word of your choosing, making the total 5 words

Feedback

Please provide feedback with this anonymous feedback form!!

https://tinyurl.com/fb-student-fa24