

Films,

Fruits, and

folklore



Introducing Ideas of Machine
Learning to High School
Students



MCTM 2024

WHO ARE WE?



CHELSEY

Our boss and
LASER Lab
webmaster



REGINA

Minister of style and
design



PABLO

Crack research and
animation ace



ANDY

Just here for the ride

Learn more at: <https://laser-umn.github.io/about.html>

Machine Learning

- Machine learning (ML) uses algorithms that “learn” from data to make predictions.
- One common application of ML is to **classify cases based on how similar they are**. This is used for:
 - Making recommendations (e.g., movies, things you might like to buy)
 - Image classification
 - Fraud detection
 - Spam filtering
- Methods used in practice can be quite complex
 - Today we will focus on foundational ideas underlying similarity quantification and classifying cases.

Classifying Movies

Introduction to Classification



MYSTERY MOVIE CHARACTERISTICS

Discuss what genre you believe this movie falls under (e.g. horror, comedy, drama, musical, action, romance, etc.)

Based on a book?	Yes
Rotten Tomatoes Score	>85%
Pass the Bechdel test?	No



MYSTERY MOVIE CHARACTERISTICS

Does your answer change with new information?

Based on a book?	Yes
Rotten Tomatoes Score	>85%
Pass the Bechdel test?	No
Musical Adaptation?	Yes (musical adaptation of the film was created)
Runtime	<120 minutes
# of Academy Award Nominations	2



Genre: Comedy



FOLLOW ALONG

Use this QR code and/or link to access the slides and worksheets we'll be using today:

<http://z.umn.edu/MCTM>



LET'S CLASSIFY!

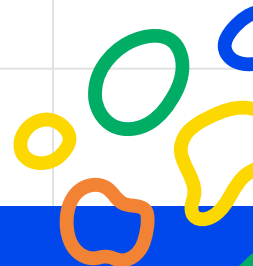
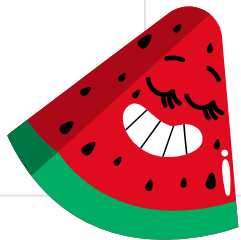
Classifying Fruits

Measures of Similarity
K Nearest Neighbors (KNN)

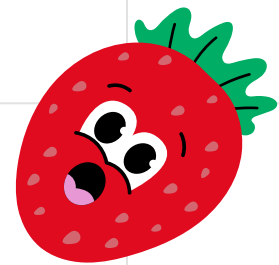
Work on:

Classifying Fruit Activity 1

(groups of 2-3)



Classifying the Goldy Fruit



- What is your method to quantify similarity?
- How did you classify the Goldy Fruit?
 - Does this line up with your first guess in #4



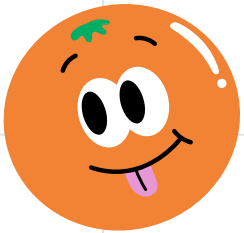
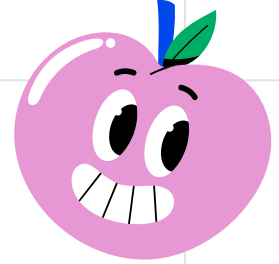
Work on:
Classifying Fruit Activity 2
(groups of 2-3)

***check in after you finish #1**



Wrap Up Discussion

- Results
- Choice of k
- Ties

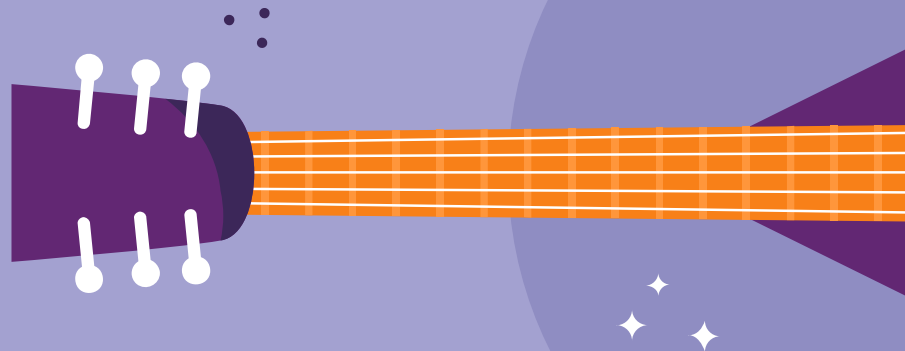




CLASSIFYING TAYLOR SWIFT

Euclidean Distance
Multiple Variables

Work on:
Taylor Swift
Activity 3
(in groups of 2-3)



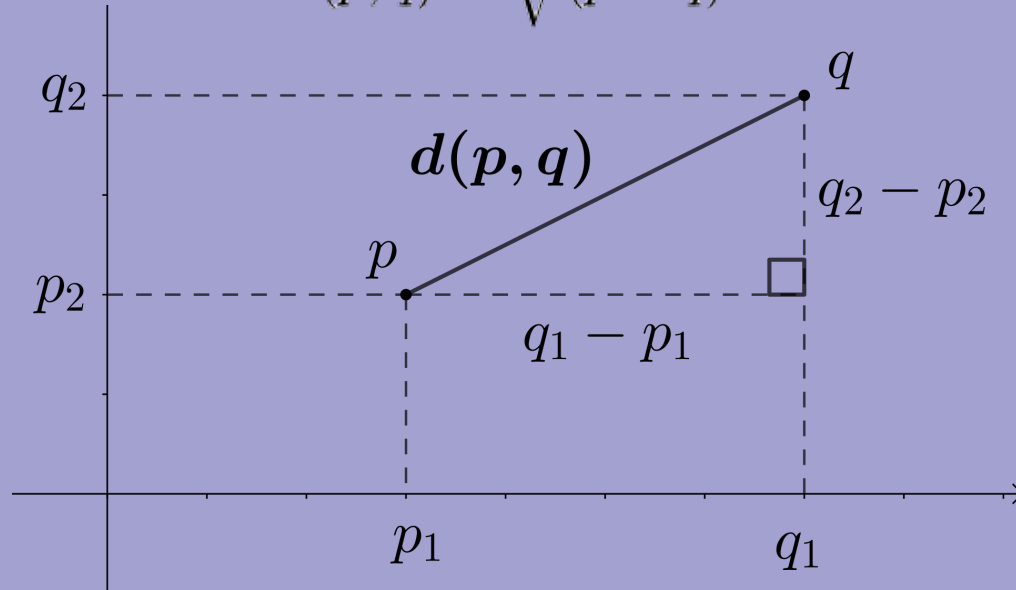


**How did you quantify
similarity between
songs?**



EUCLIDEAN DISTANCE

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2}$$

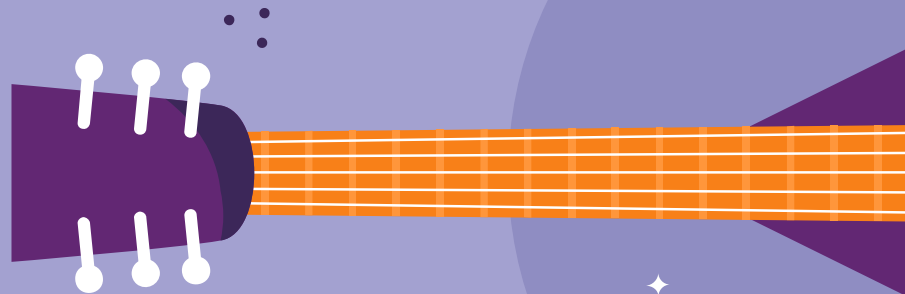




**Which album should
Tay-Tay add End of
the Road to?**



Work on:
Taylor Swift
Activity 4
(in groups of 2-3)



What album should
End of the Road be
released on (using
the optimal k)?



PROPERTIES OF SONGS

LOUDNESS

TEMPO

DANCEABILITY

ACOUSTICNESS

ENERGY

VALENCE

LIVENESS

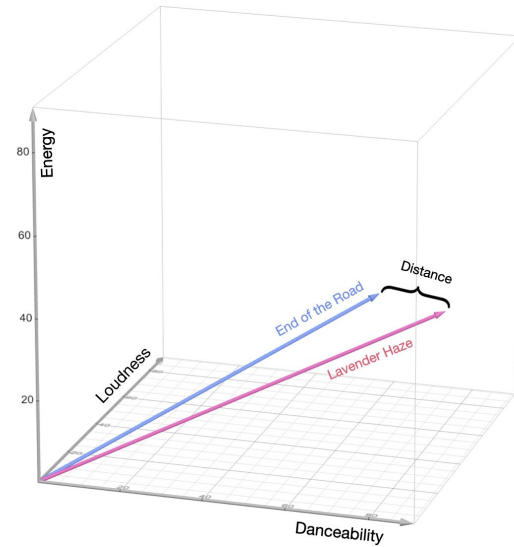
INSTRUMENTALNESS

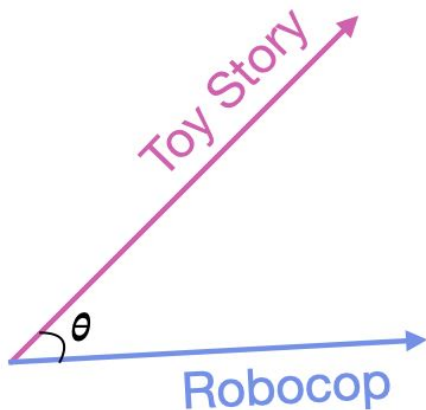
SPEECHINESS



EXTENSION OF EUCLIDEAN DISTANCE

- Euclidean distance in multidimensional space
- Vectors
- Euclidean distance via matrix algebra [Optional]





COSINE SIMILARITY

- Introduction to cosine similarity
- Movie recommendations (common application)

SIMILARITY WITH CATEGORICAL ATTRIBUTES

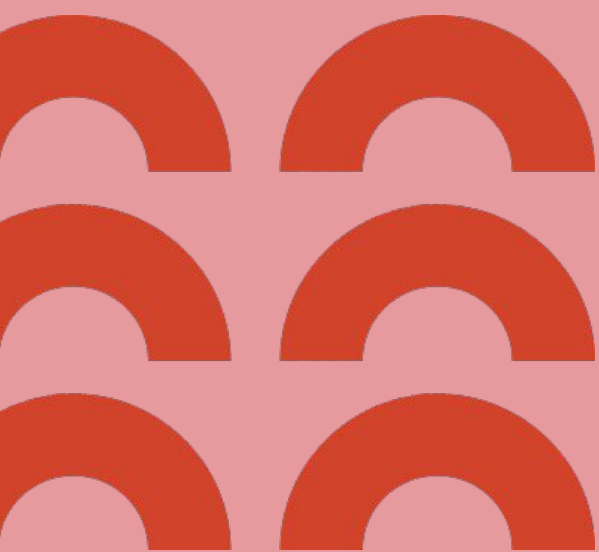


- Introduction to binary attributes
 - Symmetric binary attributes
 - Asymmetric binary attributes
- Measures for quantifying similarity between cases with categorical attributes

SIMILARITY WITH MIXED ATTRIBUTES

- Introduction to Gower's distance
- k NN when classes are imbalanced





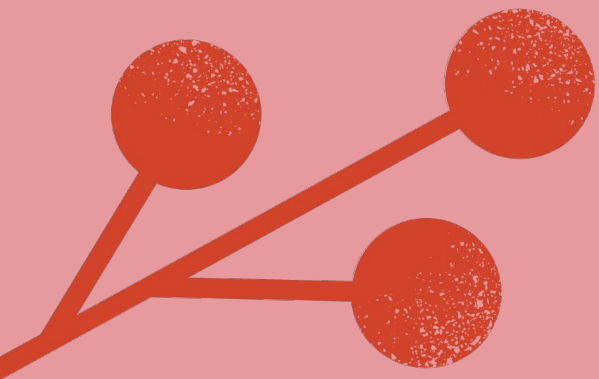
STAY CONNECTED

KEEP UP WITH OUR WORK!

- Algorithmic Modeling (there's more!)
 - Data to Graphs
- Statistics Teaching Inventory
 - Code Review



Click on the logo for our website homepage!



Be Above Average: Teaching Introductory Statistics for Deeper Understanding

Check out our other workshop!

We'll be talking about modeling and simulation activities for high school classrooms.

**SATURDAY 9:25 AM - 10:55 AM in
Gooseberry Falls 1**



THANK



YOU