Classify This!

Adventures in Algorithmic Modeling

USCOTS
2025

FOLLOW ALONG

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

http://z.umn.edu/USCOTS

LET'S CLASSIFY!

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

WHO ARE YOU?

Please share with all of us:

1. Who are you?
2. What do you teach?
3. What drew you to this session?

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

Machine Learning

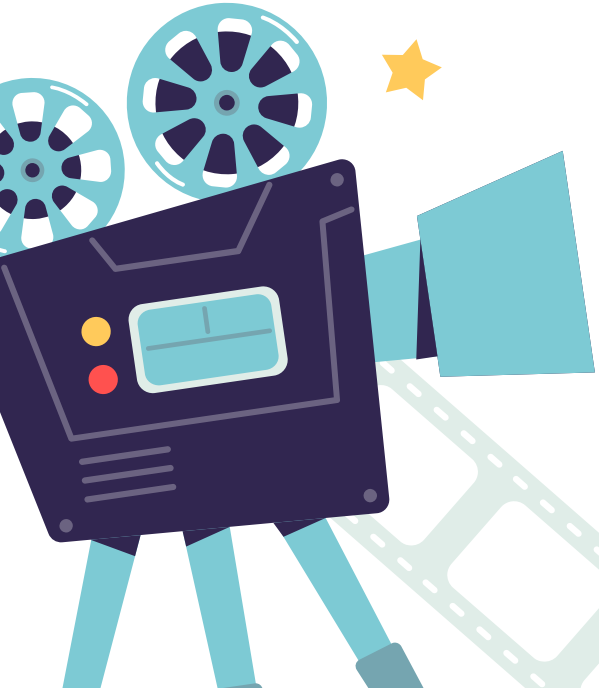
- Machine learning (ML) uses algorithms that “learn” from data to make predictions.

- 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

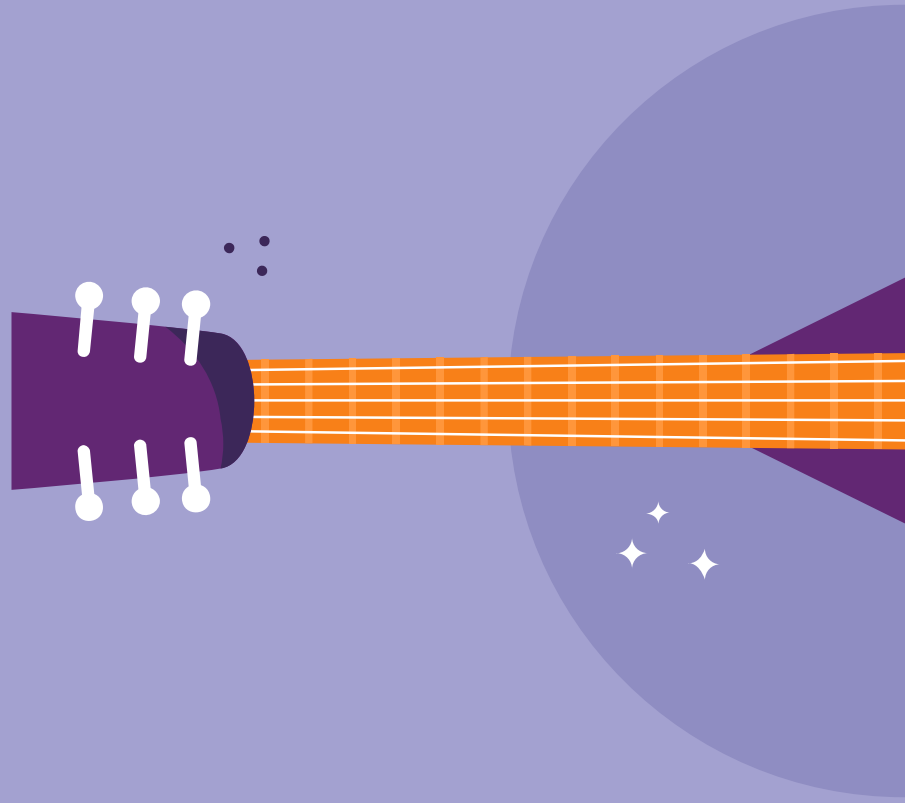
Does your answer change with new information?

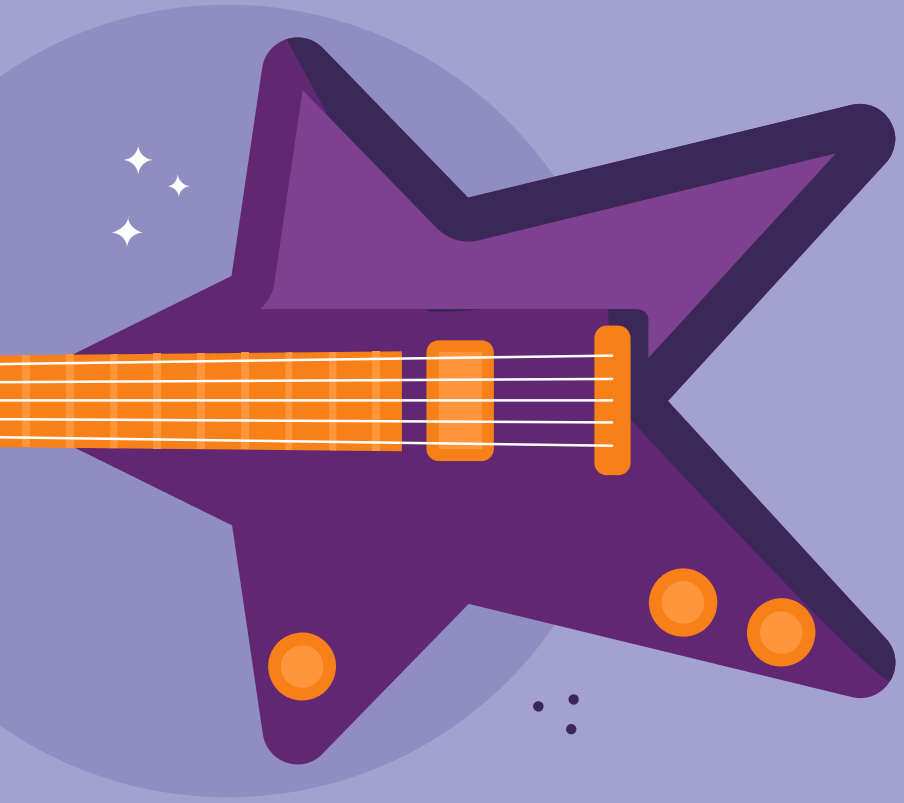
Based on a book?	Yes
Rotten Tomatoes Score	>85%
Pass the Bechdel test?	No
Musical Adaptation?	Yes (musical adaptation of the film was created)
Runtime	<120 minutes
# of Academy Award Nominations	2

CLASSIFYING TAYLOR SWIFT

Purple Rain (Taylor's Version)

(in groups of 2-3)





How did you quantify similarity between songs?
#3 and #4

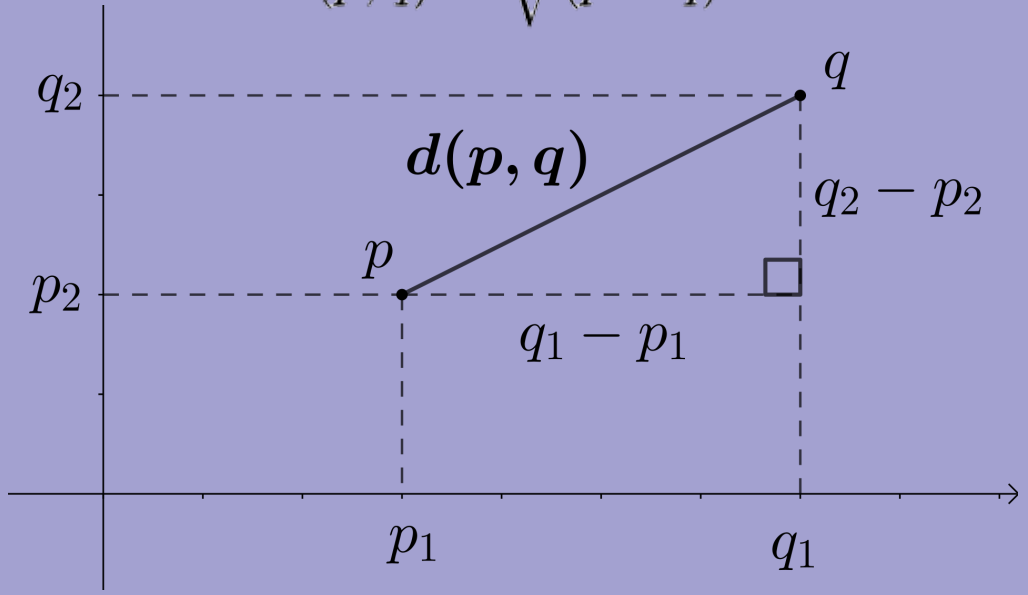
Measures of similarity

->

Euclidean distance

EUCLIDEAN DISTANCE

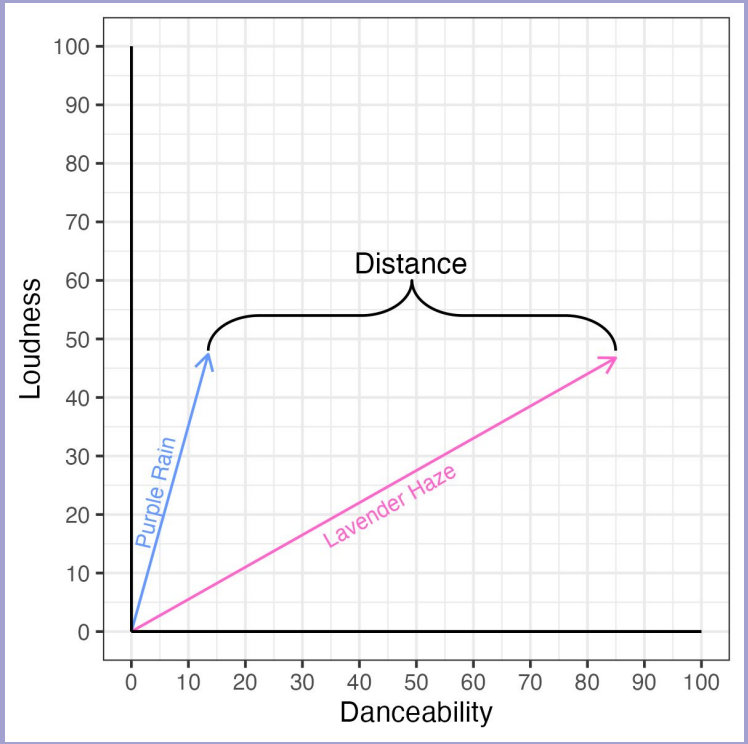
$$d(p, q) = \sqrt{(p - q)^2}$$



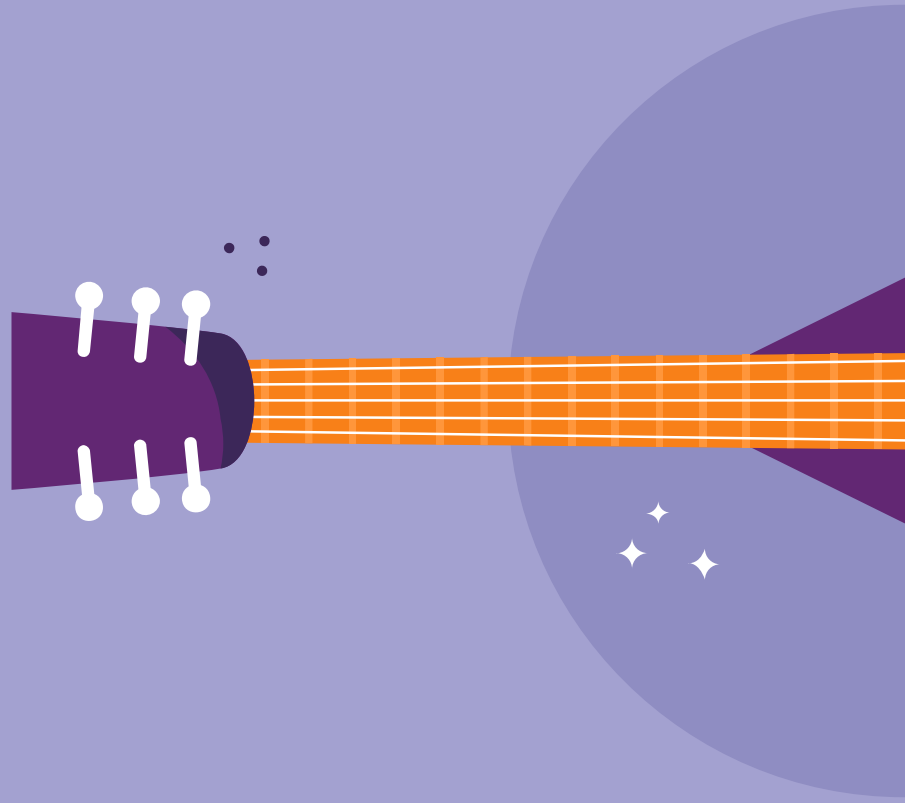
EUCLIDEAN DISTANCE

$$d(a,b) = \sqrt{\sum_{i=1}^{p} \left(a_i - b_i\right)^2}$$

EUCLIDEAN DISTANCE



Work on You Belong with Me: Classifying Taylor Swift (in groups of 2-3)



USING TECHNOLOGY

Link to Google Sheets

+

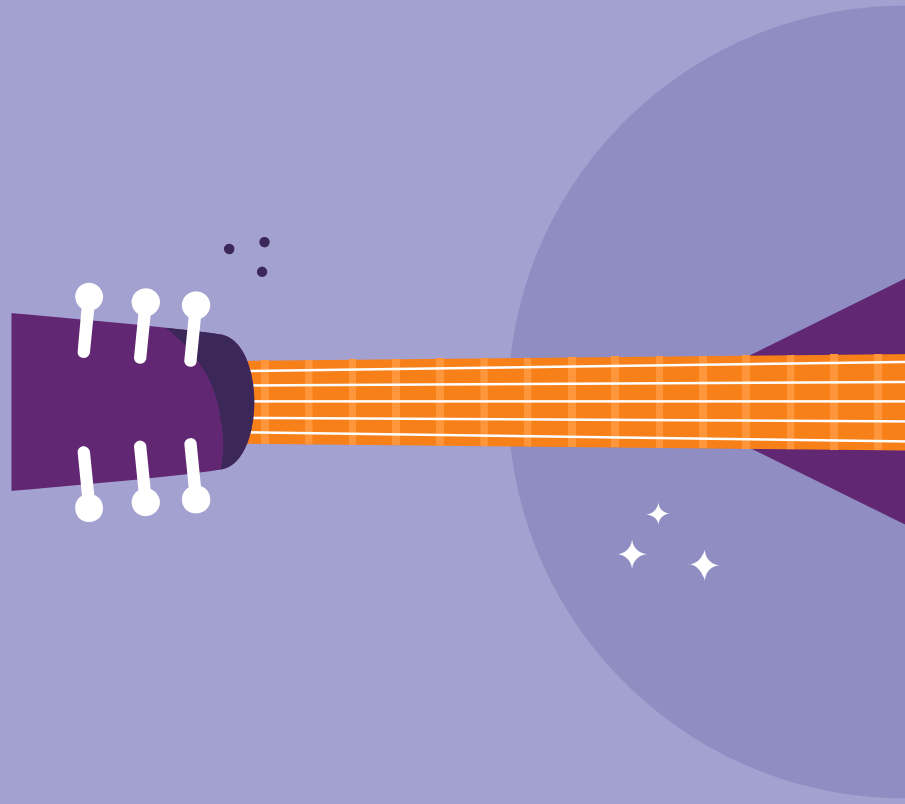
Intro to formulas

Which album should Tay-Tay add Purple Rain to?

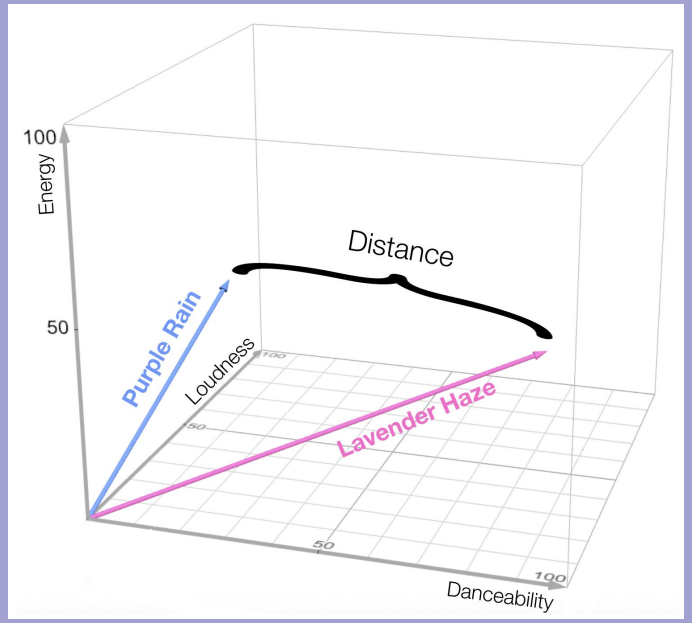
PROPERTIES OF SONGS

- LOUDNESS
- TEMPO
- DANCEABILITY
- ACOUSTICNESS
- ENERGY
- VALENCE
- LIVENESS
- INSTRUMENTALNESS
- SPEECHINESS

The Mathematics of Euclidean Distance - More than 2 variables

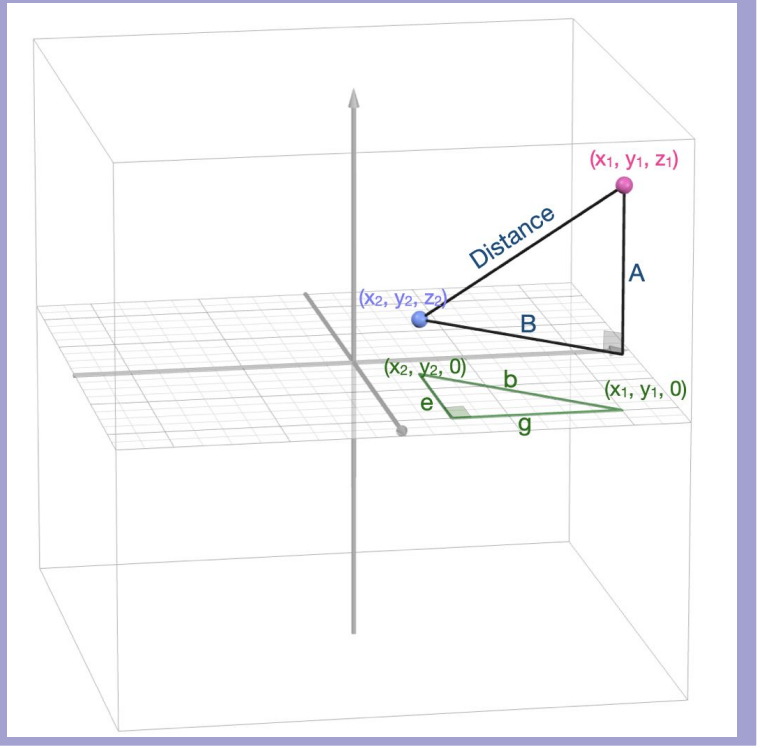


EUCLIDEAN DISTANCE IN MULTIDIMENSIONAL SPACE



PROJECTIONS

- Find the length of b in the green triangle
- b is the projection of B so b and B have the same length
- Use B and A to determine the hypotenuse of the black triangle



EXTENSIONS

- Mathematics of Euclidean Distance
- Blank (Vector) Space
- kNN

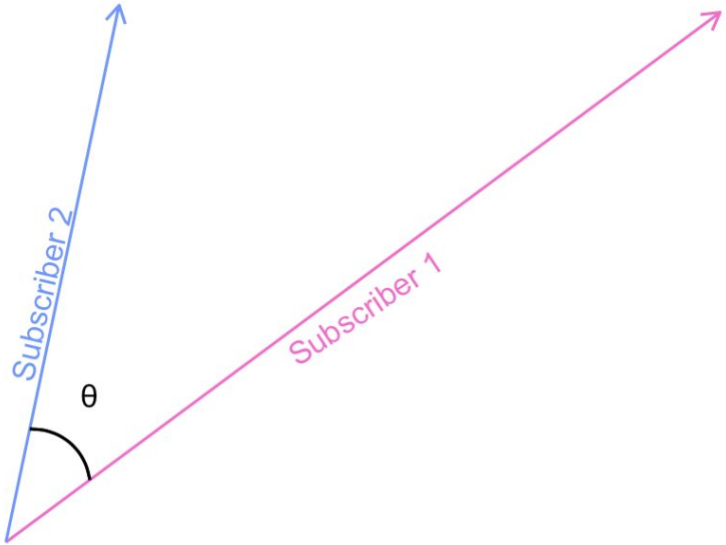
Classifying Subscribers

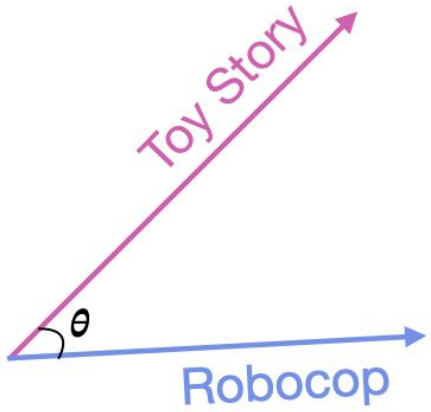
MOVIE RECOMMENDATIONS

Which two subscribers are most similar in their viewing profiles?

- **Subscriber 1:** Watched 3 Dramas, 1 Rom-Com, and 1 Action movie
- **Subscriber 2:** Watched 3 Dramas, 1 Documentary, and 1 Horror film
- **Subscriber 3:** Watched 9 Dramas, 3 Rom-Coms, and 3 Action movies

COSINE SIMILARITY





ITEM-BASED COLLABORATIVE FILTERING

User Ratings

Title	A	B	C	D	E	F	G	H	I	J
Toy Story	2.5	0	4.5	4	4	3.5	4	5	4	3
RoboCop	2	0.5	4	2.5	3.5	0	0	5	3	0

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

Discussion

- Q+A
- Where does this fit in the curriculum?
- Reminder that the suite of activities starts with kNN

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!

https://laser-umn.github.io/

THANK YOU