



# **CIS 635 Knowledge Discovery & Data Mining**

Introduction to Linear Algebra



# Survey results and project groups

- Survey summary
- Final project groups formation:
  - Each group will be comprised of two (2) students
  - You are welcome to form your own group, or we will assign groups based on your survey response (background, area of interest, and programming efficiency). Please respond to the survey if you haven't already.
  - If you want to form your group, please let us know by: **09/04/2024**. The point of contact on the following TA:  
**Sridevi Bommidi** (bommidis@mail.gvsu.edu)



# Outline

- Proximity vs Distance Metric
- k-NN, our first ML model
- Concept of Vectors and Vector operations
- Digital data, their encodings, and their representations through Vectors
- NumPy basics

# Digital data

- In computing everything is digital and binary
- All data types we talked about
- Bit( 0 / 1): Digital letter
- Byte ( 000 0011): Digital word
- Kilo (Byte), Mega(Byte), Giga (Byte): We are talking about Digital data and their sizes mainly

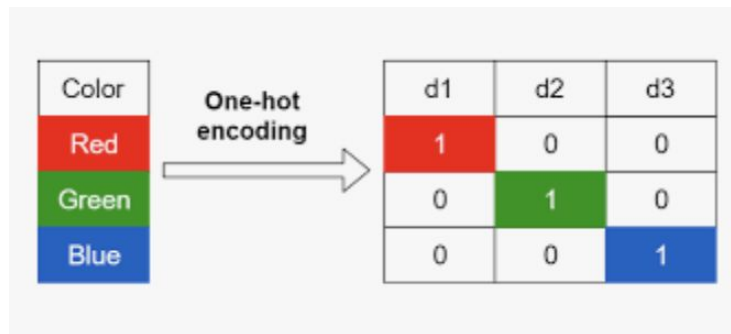




# Categorical Data

# One hot encoding

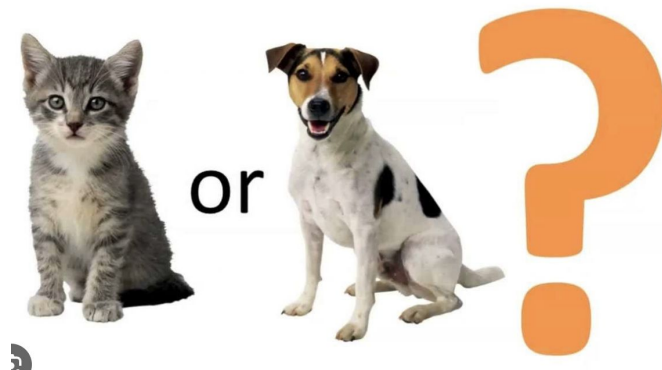
- Only one bit is 1
- A vector representation of categorical values



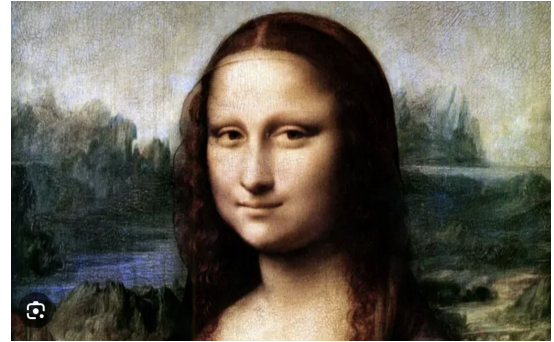
## One hot encoding (cont.)

Classification task:

- Binary example {Cat vs Dog}
- Set size is 2
  - Cat (0, 1)
  - Dog (1, 0)
  - Or vice versa
- Same rule applies every categorical data

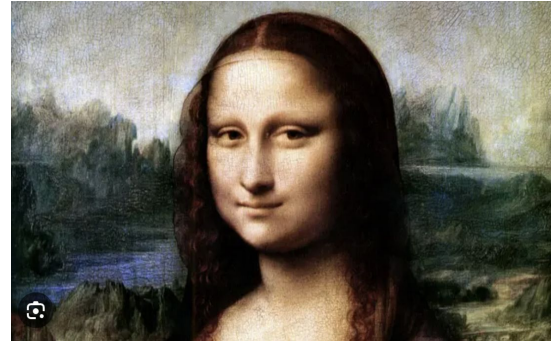


# Binary, gray-scale, and color images

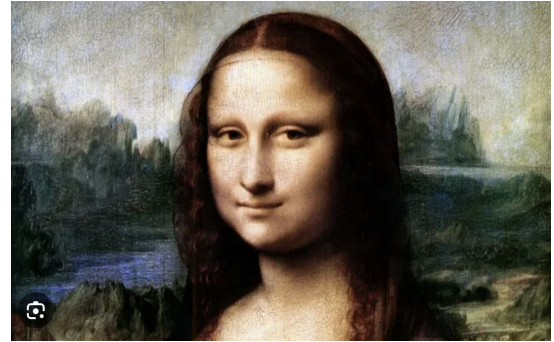
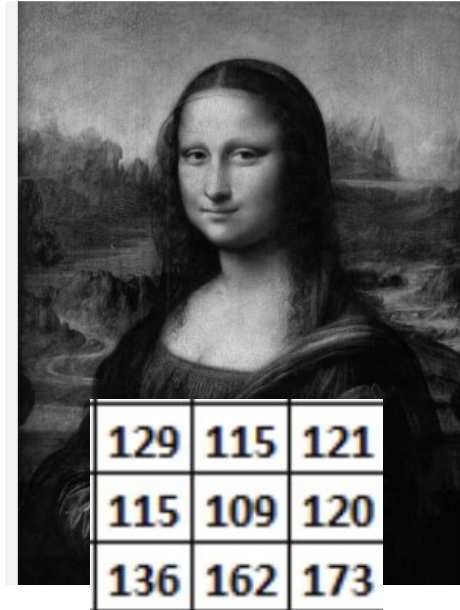




# Binary, gray-scale, and color images

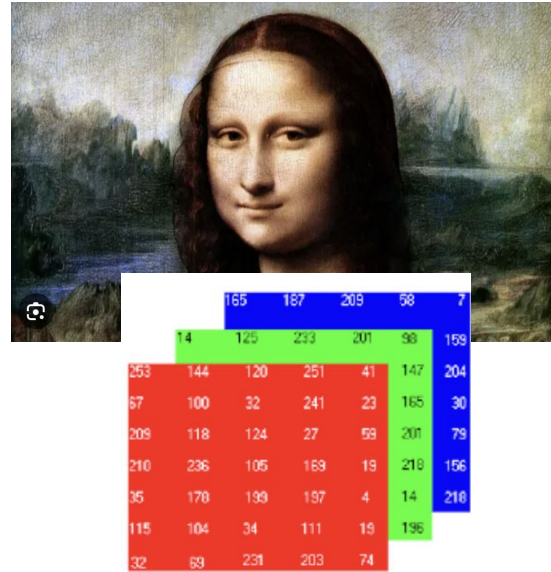
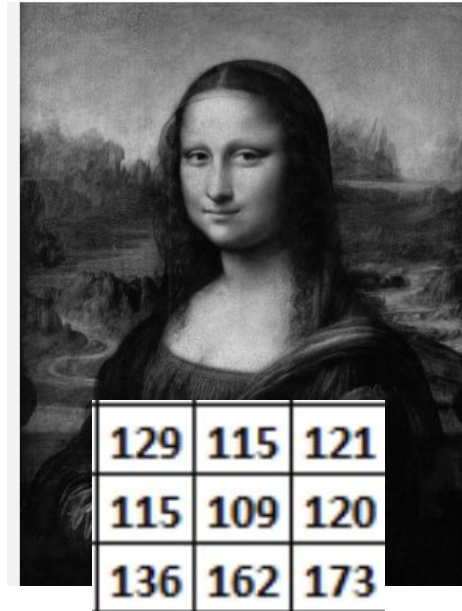


# Binary, gray-scale, and color images



Value range: [0 - 255]

# Binary, gray-scale, and color images





## Question

- What will be the **vector** size of a 40x50 RGB color image?



## Question

- What will be the **vector** size of a 40x50 **RGB** color image?
- Answer: **6000**



## Question

- How do you represent your data (some call feature encoding) for a spam filtering model, say k-NN spam filter?



# Numpy

[Let's practice](#)