Worksheet 03

Name: William Seong UID: U76365580

Topics

- Intro to DS

Linear Algebra Review

If you need a linear algebra review, please read through the following pdf (https://github.com/gallettilance/CS506-Spring2023/raw/main/worksheets/lecture_03_linear_algebra_review.pdf) before next class

Intro to Data Science

a) what property must a hypothesis have?

Must be testable.

b) what examples would you have wanted to try?

(1,2,3) and (6,4,2)

c) Poll 1

C

d) Given the hypothesis (x, 2x, 3x), for each of the following, determine whether they are positive or negative examples:

- (2, 4, 6)
- (6, 8 , 10)
- (1, 3, 5)

Positive
Negative
Negative

e) Poll 2

Type *Markdown* and LaTeX: α^2

f) Describe steps of a Data Science Workflow

Process Data: Work with the data, make it easier to handle, data engingeering, making data pipeline.
Explore Data: Try and understand how different items in the df vary as a function of each other.
Extract Feature: Extract the features or attributes that will better help you predict the model, what am I gonna describe my target as a function of?
Create Model: The end product, may need more tweaking.

g) Give a real world example for each of the following data types:

- record
- graph
- image
- text

Record: m-dimensional points/vectors, (name, age, balance) -> ("John", 20, 100)
Graph: Nodes connected by edges, maps
Image: Matrix of pixels, pictures
Text: List of words, corpus of document as a matrix, document and words, csv

h) Give a real world example of unsupervised learning

```
A recommender system that connects related articles and recommends them to users.
```

i) Give a real world example of supervised learning

```
Classification of tumors if they are benign or malignant.
```