Worksheet 03

Name: Rishab Sudhir

UID: U64819615

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?

A hypothesis must be falsifiable. (Come up with examples where the hypothesis fails)

b) what examples would you have wanted to try?

(14,16,18)

(1,4,7) (maybe just any increment?)

#(x,x+y,x+2y)

c) Poll 1

B

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

C

f) Describe steps of a Data Science Workflow

- Process Data
- Explore Data
- Extract Features
- Create Model

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

- record
- graph
- image
- text

- tables of attributes (eg excel for tracking expenses) (name, age, balance) -> ("John", 20, 100)
- An ancestry tree
- Pictures of dogs for a neural network
- Transcripts of books

h) Give a real world example of unsupervised learning

k-means clustering (eg clustering buckets for flight tickets)

i) Give a real world example of supervised learning

Linear Regression (eg housing price predictions)