Worksheet 03

Name: Mohit Sai Gutha UID: U48519832

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)](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?

Falsifiability : A hypothesis should be structured in a way that makes it possible to be proven false.

b) what examples would you have wanted to try?

(8,12,10);(111,113,115);(22,20,18)

c) Poll 1

C. (1,2,3)

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)

1. (2,4,6) : Positive
2. (6,8,10) : Negative
3. (1,3,5) : Negative

e) Poll 2

C. (1,2,3)

f) Describe steps of a Data Science Workflow

1. Process Data : Data is gathered, cleaned, and formatted for analysis.
2. Explore Data : The goal is to try and understand how different elements of the data vary as a function of another.
3. Extract Features : Identifying relevant features that will be used for analysis and modeling.
4. Create Model : Developing and training models based on the data to make predictions or derive insights.

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

- record
- graph
- image
- text

1. Record : Patient electroni health records containing medical history, diagnoses, and treatments.
2. Graph : Transportation networks data.
3. Image : Satellite Images for land cover classification
4. Text : News articles

h) Give a real world example of unsupervised learning

Emails similar to the mails marked as spam could be labelled as spam.

i) Give a real world example of supervised learning

Predicting whether tumor is benign or malignant based on age and tumor size.