

✓ Worksheet 03

Name: Jeong Yong Yang, Junho Son UID: U95912941, U64222022

Topics

- Intro to DS

Linear Algebra Review

If you need a linear algebra review, please read through the [following.pdf](#) before next class

Intro to Data Science

a) what property must a hypothesis have?

Hypothesis must be testable.

b) what examples would you have wanted to try?

We wanted to try (7,8,9), (-5, -3, -1), and (-4, -3, -2).

c) Poll 1

We chose 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)

(2, 4, 6) → positive

(6, 8, 10) → negative

(1, 3, 5) → negative

e) Poll 2

We chose C. (1, 2, 3)

f) Describe steps of a Data Science Workflow

The steps of Data Science workflow are:

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
-
- record: Boston University Student record → can store student's birthdate, ID number, major, phone number, address, and GPA
 - graph: Facebook → the nodes represent the different users/accounts in Facebook and the edges between them show whether the users are friends with one another
 - image: Image classification → images of dogs and cats and classifying them correctly as dogs or cats
 - text: Shakespeare plays → stores chapters, paragraphs, setting and play instructions for natural language processing (NLP)

h) Give a real world example of unsupervised learning

In natural language processing (NLP), topic modeling, the usage of LDA to discover possible topics of documents, is an example of unsupervised learning by looking at the patterns of words and the frequency of specific words.

i) Give a real world example of supervised learning

Classifying whether an email is spam or ham (not spam). Using previously saved emails that are spam or not spam (look at the patterns of word choices, email title and message, spellings, and etc.), we can classify whether the new set of emails are spam or not spam.

1 코딩을 시작하거나 AI로 코드를 생성하세요.