Audit Access Expires May 11, 2020
You lose all access to this course, including your progress, on May 11, 2020.

3. Introduction to Clustering
Introduction to Clustering

0:00 / 0:00 1.25x

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

Different Types of Machine Learning Problems

3/3 points (graded)

1. We wish to classify some news articles into three fixed categories–politics, sports, and entertainment. Among the following, which machine learning problem is this?

- (•) Classification
- ( ) Regression
- ( ) Clustering

✔

2. We wish to partiton Google News articles into three sets of similar articles. Among the following, which machine learning problem is this?

- ( ) Classification
- ( ) Regression
- (•) Clustering

✔

3. We would like to predict the expected value of tomorrow's NASDAQ index. Among the following, which machine learning problem is this?

- ( ) Classification
- (•) Regression
- ( ) Clustering

✔

Solution:

Classification and regression are supervised learning approaches that map the input to an output based on a training set of input-output pairs. Classification predicts categorized labels while regression predicts a continuous value output. Clustering is an unsupervised learning approach where we do not know the label of the data. Clustering is used to visually get a sense of how many potential groupings are there in the data. In other words, the ultimate goal of clustering is **visualization of information**.

Submit You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Counting

1/1 point (graded)

Consider the following small paragraph that we wish to cluster. We want our clustering algorithm to identify as a cluster text that gives information **about** a person.

Text: Alice commented on Bob's (a new Software Engineer hire in the team) approach to solving the problem. Alice was quoted in the meeting as saying "Bob's algorithm is theoretically optimal. But, the algorithm is very hard to implement in real time. After all, he does not have an EE background and does not yet understand the way read thresholds work in a memory chip. He needs to come up to speed with what is possible in the read channel. So, set up a meeting with the hardware and chip design engineers." Alice then ended the meeting.

Consider the following “clustering" algorithm: The algorithm assigns as the label to the text the name of the person that appears most in the text.

Does this algorithm achieve our clustering objective?

Yes

No

Solution:

Alice's name appears thrice in the given text and Bob's name appears twice. According to the algorithm, this text will be labeled "Alice". However, we can clearly infer from the text that the text provides a lot more information about Bob than it does about Alice.

Submit You have used 1 of 1 attempt

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 4 Unsupervised Learning (2 weeks) :Lecture 13. Clustering 1 / 3. Introduction to Clustering

Add a Post

Show all posts by recent activity

STAFF - Resource posted for new lectures Can you please post the handouts for the new lectures. They really help. Thanks. Can you go ...	4
Counting clustering problem Is it just me? I inferred a lot of information about Alice from this text; how she assesses work,...	4
Counting question The text does indicate that Alice is rude.	2
[STAFF] Feedback to video editing IMO The camera angle change showing the blackboard from the side is not helpful and distra...	2

[staff] Google clustering approach 3