Hide menu

Outlier Detection

Congratulations

Graded Assignment: Self Reflection
Started

Discussion Prompt: Data Analysis Project Show Off!
1h

Reading: Congratulations!
10 min

Self Reflection

Review Learning Objectives

Assignment details

Due
Mar 3, 11:59 PM IST

Attempts
Unlimited

Your grade

You haven't submitted this yet. We keep your highest score.

--

Like Dislike Report an issue

Instructions

1. Reflecting on the case study, what was the most challenging aspect of applying classification analysis to solve the real-world problem? How did you overcome this challenge, and what did you learn from it? [Practice this question as if you were in an interview!] **1 point**

The most challenging part is feature engineering, preprocess the data to compute the right set of features and then select the right classification model. A few feature selection models and feature extraction models (e.g., PCA) can be used, along with standard set of preprocessing techniques (e.g, z-score normalization, imputation for the missing values) applied prior to applying the model.

Your answer cannot be more than 10000 characters.

Resume

2. Reflecting on the case study, what was the most challenging aspect of applying regression analysis to solve the real-world problem? How did you overcome this challenge, and what did you learn from it? [Practice this question as if you were in an interview!] **1 point**

Similar to the classification case study, here also preprocessing is a major challenge for regression analysis too. For example, we need to get rid of the features not carrying much information (e.g., car_ID, CarName etc.) and then for some regression models (e.g., SVR) we need to convert the categorical features to OHE (one-hot-encoding), prior to training the model.

Your answer cannot be more than 10000 characters.

3. Reflecting on the case study, what was the most challenging aspect of applying clustering and dimension reduction analysis to solve the real-world problem? How did you overcome this challenge, and what did you learn from it? [Practice this question as if you were in an interview!] **1 point**

While using the DBSCAN algorithm, choosing the threshold (epsilon) for neighborhood is sometimes a challenge. The data contained the ground-truth labels, hence matching the cluster-labels output by the algorithm with the ground-truth, we can intuitively create most probable mapping between them and then compute the accuracy of clustering. Although Silhouette coefficient gives some idea about the quality of the clusters, whether different / same species belong to different / same cluster respectively, we can compute it using the GT labels.

Your answer cannot be more than 10000 characters.

4. Reflecting on the case study, what was the most challenging aspect of applying association rule mining and outlier detection to solve the real-world problem? How did you overcome this challenge, and what did you learn from it? [Practice this question as if you were in an interview!] **1 point**

It's the preprocessing step, again. The python implementation of the apriori and fp-growth algorithms expect the input data-structure as transaction encoder. not as pandas Dataframe or numpy darray. We need to create a TransactionEncoder object and fi t the dataset into it, using the module mlxtend.preprocessing, but it accepts a list of list instead of a Dataframe, so a prior conversion is needed.

For outlier detection, we used contaminated MNIST dataset and again choosing the right set of features from the flattened 28x28 images, along with the algorithm to be chosen for outlier detection (e.g., one-class SVM, isolation forest, LOF etc.) was itself a challenge, along with evaluation of the algorithms.

Your answer cannot be more than 10000 characters.

5. Are you ready to apply the knowledge you gained, and to change the world? **1 point**

Well, I can definitely try to apply the algorithms, but how much the world will change with that is questionable ;-)

Your answer cannot be more than 10000 characters.

Coursera Honor Code Learn more

I, **SANDIPAN DEY**, understand that submitting work that isn't my own may result in permanent failure of this course or deactivation of my Coursera account.

Submit **Save draft**

Last saved on Mar 3, 3:18 AM IST

Like Dislike Report an issue