Big Data Analytics Unit 4

David Finger

Colorado Technical University

Abstract

Explain how prediction algorithms can score new records.

K Nearest Neighbors, here within forth referred to as KNN, is a machine learning language algorithm that can be used for predicting new data points. KNN takes provided data and creates training points that it uses to determine where to classify all other data points within the set. Using this data, KNN can group data points together using the average of its neighboring data points.

When KNN scores new data points, or predicts where new data points will be, it assigns those predicted points based on where the data can fall. This is regarding where a point would have the most neighbors, when the data is being used for classification. Or where the point would be within the average of all neighbors, which is used only with regression data. These two methods allow KNN to create semi accurate predictions about where new data points may fall within the datasets.

Resources

Khandelwal, R. (2018, November 16). *K-Nearest Neighbors(KNN)*. Medium. <https://medium.com/datadriveninvestor/k-nearest-neighbors-knn-7b4bd0128da7>