

Clustering&Classification

Einführung:

Die Daten spielen eine wichtige Rolle in unserem heutigen Leben, von den alle Forschungen und Studien abhängen.

bevor man sich mit Classification and Clustering befasst, es ist notwendig, die Bedeutung der Daten zu kennen.

- Im Allgemeinen handelt es sich bei den Daten um eine Menge von Buchstaben, Wörtern, Zahlen, Symbolen oder Bildern, die sich auf ein bestimmtes Thema beziehen.

Beispiel: Studentendaten (Name, Matrikelnummer, Geburtsdatum, Studienfach, usw.).

Die Bedeutung von Clustering:

Clustering ist eine Methode, Objekte so zu gruppieren, dass Objekte mit ähnlichen Merkmalen zusammenkommen und Objekte mit unterschiedlichen Merkmalen auseinander gehen.

Es ist eine gängige Technik für die statistische Datenanalyse im maschinellen Lernen und Data Mining.

Clustering ist eine Methode zum unbeaufsichtigten Lernen, mit der ähnliche Instanzen auf Basis von Features gruppiert werden.

Die Bedeutung von Classification:

Classification ist ein Kategorisierungsprozess, bei dem Objekte anhand des Trainingssatzes von Daten erkannt, differenziert und verstanden werden.

Klassifikation ist eine Methode des überwachten Lernens, mit der Instanzen anhand von Merkmalen vordefinierte Tags zugewiesen werden.

Example:

Angenommen, Sie haben einen Korb voller verschiedenen Bällen und Ihre Aufgabe besteht darin, die Bälle, die die gleiche Größe haben, an einem Ort anzuordnen.

Angenommen, die Bälle sind Fußbälle, Basketbälle und Handbälle.

So wissen Sie bereits aus Ihrem vorherigen Wissen, dass es in Form aller Bälle einfach ist, die Bälle an einem Ort anzuordnen.

Hier wird Ihr vorheriges Wissen als trainierte Daten im Bereich Data Mining bezeichnet.

Diese Art von Daten erhalten Sie aus trainierten Daten. Diese Art des Lernens wird als überwachtes Lernen bezeichnet.

Diesmal wissen Sie gar nichts über diese Bälle, Sie sehen diese Bälle zum ersten Mal, also wie werden Sie die gleiche Art von Bällen anordnen?

Was Sie zuerst tun werden, ist, dass Sie die Bälle nehmen und Sie werden jede natürliche Eigenschaft diesen bestimmten Ball auswählen.

Angenommen Sie haben die Größe genommen. Dann werden Sie basierend auf der Größe anordnen. Dann werden die Gruppen so etwas sein.

Große Bälle Gruppe: Basketbälle.

Mittelgroße Bälle: Fußbälle.

Kleine Bälle: Handbälle.

Diese Art des Lernens wird als unbeaufsichtigtes Lernen bezeichnet.

Was ist der Unterschied zwischen Clustering und Klassifizierung?

Merkmale von Clustering und Klassifizierung:

Supervision:

Clustering: Clustering ist eine unbeaufsichtigte Lernmethode.

Klassifikation: Klassifikation ist eine überwachte Lernmethode.

Klassifikation ist eine überwachte Lernmethode

Sie geben als Input charakteristische Beispieldaten mit korrekten Antworten an. Sie werden von diesem Algorithmus lernen und beginnen, die korrekten Ergebnisse basierend auf der Eingabe danach vorherzusagen. Beispiel: Spamfilter

: Clustering ist eine unbeaufsichtigte Lernmethode.

Supervision Learning: Sie geben nur Daten und sagen nichts - wie korrekte Bewertungen oder korrekten Antworten. Der Algorithmus analysiert automatisch Datenmuster. Beispiel: Google News

Trainingsset:

Clustering: Beim Clustering wird kein Trainingssatz verwendet.

Klassifikation: Ein Trainingssatz wird verwendet, um Ähnlichkeiten in der Klassifikation zu finden.

Prozess:

Clustering: Statistische Konzepte werden verwendet und Datensätze werden in Untergruppen mit ähnlichen Merkmalen aufgeteilt.

Klassifikation: Klassifikation verwendet die Algorithmen, um die neuen Daten gemäß den Beobachtungen des Trainingssatzes zu kategorisieren.

Labels:

Clustering: Im Cluster sind keine Labels enthalten.

Klassifizierung: Für einige Punkte gibt es Etiketten (Labels)

.

Ziel:

Clustering: Das Ziel von Clustering besteht darin, eine Gruppe von Objekten zu gruppieren, um herauszufinden, ob eine Beziehung zwischen ihnen besteht.

Klassifikation: Ziel von Klassifikation ist es, herauszufinden, zu welcher Klasse ein neues Objekt gehört, aus dem Satz vordefinierter Klassen.

Combination von Clustering und Classification:

Clustering ist nicht nur ein unbeaufsichtigtes maschinelles Lernen, sondern kann auch dazu verwendet werden, Cluster als Features zur Verbesserung von Klassifizierungsmodellen zu erstellen. Sie alleine reichen nicht für eine Klassifizierung, wie die Ergebnisse zeigen. Bei Verwendung als Features verbessern sie jedoch die Modellgenauigkeit

When to apply them?

Wir brauchen Clustering in :

Identifizierung der Konzertzelle

Such Maschinen (Google)

Wireless Network

Kundensegmentierung (Customer Segmentation)

Wir brauchen Classification in :

Identifizierung der Konzertzelle

Email spam

Erkennung der wichtigsten Gesichtspunkte (Facial key points detection)

Drugs Classification

<https://data-flair.training/blogs/clustering-in-machine-learning/>
Beispiele

Source:

https://www.researchgate.net/figure/An-example-of-the-document-clustering_fig1_322455242 Foto von Clustering

<https://blogs.sas.com/content/subconsciousmusings/2016/05/26/data-mining-clustering/>

Foto clustring : uterscheiden von Miterareiten von verschiedene Berufe

<https://im-coder.com/unterschied-zwischen-klassifizierung-und-clustering-im-data-mining.html>