Evaluation

Daten Vorbereitung

Für jede Methode werden die Daten auf unterschiedliche Weise aufbereitet. Für die auf neuronalen Netzen basierenden Methoden werden die Daten mit den folgenden Skalierern aufbereitet: kein Skalierer, MinMaxScaler, StandardScaler, BooleanScaler.

Bei Methoden, die auf Clustern basieren, werden die Daten mit mehr Skalierern aufbereitet, da sie einfacher und schneller mit diesen Methoden zu vergleichen sind, die den Vorteil haben, dass sie keine zeitaufwändigen Lernphasen durchlaufen.

Keine Skalierung bedeutet, dass die Daten nicht vorbereitet werden und daher direkt den Methoden zugewiesen werden.

MinMaxScaler: Skaliert den Datensatz neu, sodass alle Feature-Werte im Bereich [0, 1] liegen.

Standard Scaler: Entfernt den Mittelwert und skaliert die Daten auf Einheitsvarianz.

MaxAbsScaler: Skaliert so, dass die Trainingsdaten im Bereich [-1,1] liegen, indem jeder Datenpunkt durch den Maximalwert geteilt wird.

RobustScaler: Die Zentrierungs- und Skalierungsstatistik vom RobustScaler basiert auf Perzentilen und wird daher nicht von einigen wenigen sehr großen Randausreißern beeinflusst.

PowerTransformer: Wendet eine Leistungstransformation auf jedes Feature an, um die Daten Gauß-ähnlicher zu machen, um die Varianz zu stabilisieren und die Schiefe zu minimieren. Derzeit werden die Yeo-Johnson- und Box-Cox-Transformationen unterstützt und der optimale Skalierungsfaktor wird bei beiden Methoden über die Maximum-Likelihood-Schätzung bestimmt.

QuantileTransformer: Wendet eine nichtlineare Transformation an, sodass die Wahrscheinlichkeitsdichtefunktion jedes Merkmals auf eine gleichmäßige oder Gaußsche Verteilung abgebildet wird.

Normalizer: Skaliert den Vektor für jede Probe neu, um unabhängig von der Verteilung der Proben eine Einheitsnorm zu erhalten.

Training

Ergebnis