

Big Data / Data Mining

Künstliche Neuronale Netze (KNN)

Aufgabe 1

Lesen Sie die Seiten 117 bis 122 in Cleve und Lämmel (2014).

- Für welche Data-Mining-Aufgaben sind KNN geeignet?
- Erläutern Sie, wie Sie die Anzahl der Neuronen in der Ausgabe-Schicht bestimmen können. Welche Information kann man in der Ausgabe-Schicht ablesen?
- Was hat maßgeblichen Einfluss auf die Anzahl der Neuronen in der Eingabe-Schicht?
- Wie können Sie mit reelwertigen Attributen umgehen?
- Wie ermitteln Sie die Anzahl der Zwischenschichten sowie die Anzahl deren Neuronen?
- Erläutern Sie kurz den Begriff *Backpropagation of error*.

Aufgabe 2

Restaurant-Daten mit RapidMiner

Erstellen Sie einen RapidMiner-Prozess und laden Sie die Restaurant-Daten aus der CSV-Datei. Erstellen Sie ein KNN und klassifizieren Sie die letzten drei Datensätze. Vergleichen Sie Ihre Ergebnisse mit denen aus dem Praktikum Entscheidungsbaum II.

- Überlegen Sie sich zunächst, wie Sie Trainings- und Testmenge trennen müssen.
- Nutzen Sie den Operator *Neural Net* zur Erstellung eines KNN und wenden Sie das Modell auf Ihre Daten an. Was stellen Sie fest? Was können Sie dagegen tun (vgl. Abbildung 1)?
- Beobachten Sie, wie sich die Anzahl an Eingangs- und Ausgangsneuronen mit unterschiedlichen Codierungen der Nominalen Daten verändert.

Aufgabe 3

Iris-Daten mit RapidMiner

Erstellen Sie einen RapidMiner-Prozess und laden Sie die Iris-Daten aus dem Repository.

- Klassifizieren Sie die Daten mittels KNN (Neural Net).
- Wenden Sie die Kreuzvalidierung an.
- Optimieren Sie Ihr Modell, indem Sie verschiedene Werte für die Anzahl an Trainingsläufen ausprobieren (automatisiert).

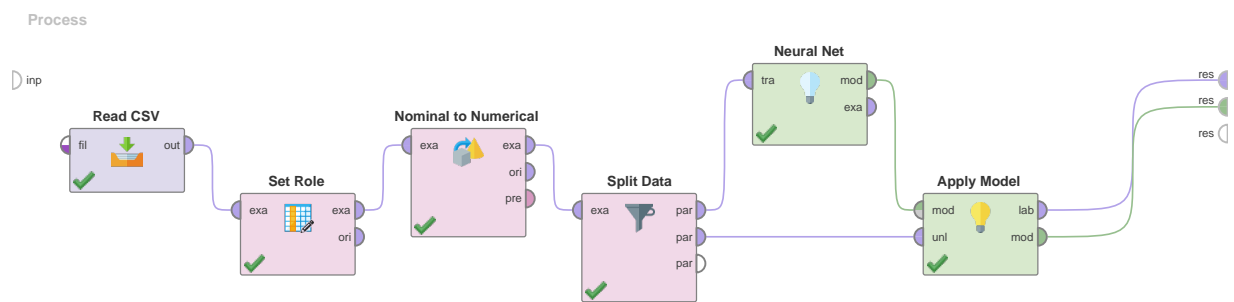


Abbildung 1: RapidMiner-Prozess: KNN mit Restaurant-Daten

Literatur

Cleve, Jürgen und Uwe Lämmel (2014). *Data mining*. Studium. München: De Gruyter Oldenbourg. 306 S. ISBN: 978-3-486-72034-1 978-3-486-71391-6.