



Statistik III: Multivariate Stat. Verfahren

Übungsblatt 9

Bearbeitung: Do. 15.7.2004, 16.00 Uhr.

Alle Aufgaben können im Mac-Cip Pool bearbeitet werden.

1. Wir betrachten den **Crabs** Datensatz.
 - (a) Erstellen Sie mit der Funktion `nnet` aus der `nnet` Library eine Klassifikation durch ein Neuronales Netz mit einem Neuron im Hidden Layer für die Variable `Sex`.
 - (b) Vergleichen Sie die Ergebnisse für Netze mit und ohne Skip-Layer.
 - (c) Visualisieren Sie für ein beliebiges Paar von Variablen die Entscheidungsregionen für je eine "gute" und "schlechte" Lösung.
2. Erstellen Sie ein Neuronales Netz für die XOR-Funktion mittels Neuronen im hidden Layer und/oder direkten Skip-Layer Verbindungen (Hinweis: Überlegen Sie zunächst, wie man die XOR-Funktion durch OR und AND abbilden kann).
 - (a) Bestimmen Sie Gewichte für Ihr vorgeschlagenes Netz, welche die XOR-Funktion abbildet.
 - (b) Generieren Sie Daten wie auf Seite 224, und passen Sie mittels der Funktion `nnet` ein Neuronales Netz mit dem von Ihnen vorgeschlagenen Layout an.
 - (c) Visualisieren Sie die Regionen, und vergleichen Sie die Gewichte des Netzes mit Ihren Gewichten aus (a).
3. Berechnen Sie für alle Fettsäuren aus dem **Olive Oil** Datensatz für die Trainingsdaten (Variable `test/training`) ein Neuronales Netz mit linearem Output, 3 Neuronen im Hidden Layer und Skip Layer Gewichten zur Vorhersage der Region.

Testen Sie die Qualität der Vorhersage des Netzes für verschiedene Konvergenz-Ergebnisse auf den Test Daten, und vergleichen Sie diese mit dem jeweiligen Ergebnis auf den Trainingsdaten.
4. Wir betrachten den **Crabs** Datensatz.
 - (a) Teilen Sie den Datensatz sinnvoll in Testdaten und Trainingsdaten auf.
 - (b) Berechnen Sie für $k = 1, \dots, 50$ je ein Neuronales Netz zur Vorhersage der vier Gruppen mit k Neuronen im Hidden Layer auf den Trainingsdaten.
 - (c) Plotten Sie je eine Reihe für:
 - i. den Vorhersagefehler der Trainingsdaten
 - ii. den Vorhersagefehler der Testdaten
 - iii. den Quotienten aus dem Vorhersagefehler der Testdaten und der Trainingsdaten.
5. Berechnen Sie analog zu der Vorgehensweise auf Seite 233 der Vorlesung eine 10-fold Cross-Validation der Vorhersage des Neuronales Netzes der **Olive Oil** Daten. Benutzen Sie als Komplexitätsparameter die Anzahl Neuronen im Hidden Layer. Für welche Anzahl von Neuronen wird der Wert der CV minimiert?

Verwenden Sie dabei alle Fettsäuren, linearen Output und Skip Layer Verbindungen.