

Übungen zur Vorlesung „Maschinelles Lernen und Künstliche Intelligenz aus Sicht der Stochastik“

Sommersemester 2017, Blatt 4

Abgabetermin: 23.05.2017, spätestens zu Beginn der Vorlesung

(Bitte geben Sie auf jedem Lösungsblatt Ihren Namen an)

Bitte maximal zu zweit abgeben!

Bearbeiten Sie R-Aufgaben grundsätzlich mit „R Markdown“ und senden Sie sie an maschinelleslernen17@gmail.com.

Aufgabe 1

(4 Punkte)

Erstellen Sie anhand der Titanikdaten `data(ptitanic)` aus dem Paket `rpart.plot` einen Klassifikationsbaum für die Zielvariable `survived` mit allen verfügbaren Prädiktoren.

- Lassen Sie den Baum wachsen und kürzen Sie ihn durch Pruning.
- Bestimmen Sie die Tiefe des Baumes bei der der Kreuzvalidierungsfehler minimiert wird.
- Vergleichen Sie die Ergebnisse für den Gini Index und die Entropy als Splitting-Kriterium.
- Von den 1309 Passagieren im Datensatz haben 500 überlebt und 809 nicht. Welchen Einfluß hat dies auf die bedingten Vorhersagefehler. Was ändert sich wenn Sie die prior Wahrscheinlichkeiten über den Parameter `parms` in `rpart` anpassen?

Aufgabe 2

(4 Punkte)

Betrachten Sie erneut das Datenset aus Aufgabe 3 von Blatt 2. Erstellen Sie erneut eine Klassifizierung mit Hilfe von Support Vector Machines.

- Benutzen Sie dafür mindestens drei verschiedene Kernel um die Punkte zu klassifizieren und entscheiden Sie sich für den geeignetesten Kernel.
- Visualisieren Sie die Entscheidungsgrenzen der Support Vector Machines.
- Begründen Sie Ihre Wahl sowohl anhand der Daten als auch anhand des Kernels.