

Maschinelles Lernen und Data Mining
Sommersemester 2008
Übungsblatt 4

Besprechung des Übungsblattes am 09.06.2008

Aufgabe 4-1 Lineare Regression
schriftlich bearbeiten

Gegeben die folgenden Daten der Modellvariable X und deren Ausprägungen Y:

x	23	24	25	26	27	28
y	150	155	150	170	160	175

- a) Nehmen Sie an, dass das Modell folgenden linearen Zusammenhang aufweist:
 $y_i = a + \beta x_i = x^T w$
Bestimmen Sie w mit Hilfe des LS-Schätzers aus der Vorlesung.
- b) Nehmen Sie nun den nichtlinearen Zusammenhang
 $y_i = a + \beta_1 x_i + \beta_2 x_i^2 = x^T w$
an und bestimmen Sie erneut w
- c) Wie könnte man den empirischen quadratischen Fehler zwischen Modell und Daten grafisch notieren?
Erklären und skizzieren Sie ihren Vorschlag im zwei- und dreidimensionalen Datenraum mit beliebigen Daten.

Hinweis: Matrixoperationen (invertieren etc.) müssen nicht von Hand berechnet werden. Im CIP-Pool steht z.B. Maple zur Verfügung (Aufruf via xmaple).

Aufgabe 4-2 Regularisierung / Overfitting
schriftlich bearbeiten

- a) Was versteht man unter dem Begriff Overfitting und wie kommt es zustande.
- b) Wie kann man erkennen, dass ein Modell „overfitted“ ist?
- c) Wie kann man Overfitting verhindern?