

Maschinelles Lernen

Übungsblatt 4

Prof. Dr. Dr. Lars Schmidt-Thieme, Osman Akcatepe
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Universität Hildesheim

29. November 2011
Abgabe: 5. Dezember bis 14.15

Aufgabe 1: IRLS (5 Punkte)

Seien folgende Daten gegeben:

y	x	y	x
0	9.5	1	11.1
0	9.6	1	11.1
0	9.7	1	11.1
0	9.8	1	11.5
0	9.9	1	11.8
0	10.5	1	11.9
0	11.0	1	12.1
0	11.2	1	12.2
0	11.5	1	12.5
0	11.7	1	12.6
0	12.1	1	12.6

a)

Wenden Sie (in R) die lineare Regression für die Zielvariable y an. Berechnen Sie die den durchschnittlichen quadratischen Fehler des Modells auf den gegebenen Daten.

b)

Berechnen Sie ein logistisches Regressionsmodell für die Zielvariable y . Wenden Sie (auf Papier, die auftretenden Gleichungssystem können Sie mit dem Rechner lösen) hierzu den Algorithmus *iteratively reweighted least squares* an. Stoppen Sie nach der zweiten Iteration. Berechnen Sie die den durchschnittlichen quadratischen Fehler des Modells auf den gegebenen Daten.

c)

Plotten Sie die Daten sowie die geschätzten Funktionen aus (a) und (b). Diskutieren Sie die Ergebnisse.

Aufgabe 2: Statistische Modelle in R (3 Punkte)

Lesen Sie Kapitel 11 von „An Introduction to R“.

a)

Mit welcher Formel würde man eine lineare Regression ohne Intercept ausdrücken? Geben Sie eine Formel für eine (nicht-triviale) polynomielle Regression an.

b)

Wie macht man Vorhersagen mit einem mittels `lm()` berechneten Modell?

Aufgabe 3: Lineare und logistische Regression in Weka (2 Punkte)

Führen Sie mit einem Regressions- und einem Klassifikationsdatensatz Ihrer Wahl in Weka eine lineare bzw. logistische Regression durch.

a)

Welche der angezeigten Fehlermaße sind sinnvoll, um ein Regressions- bzw. Klassifikationsmodell zu beurteilen? Nennen Sie jeweils ein Maß und begründen Sie kurz.

b)

Spielen Sie in beiden Fällen mit dem `ridge`-Parameter und überprüfen Sie, ob dies zu besseren Ergebnissen führt.