

A geometric approach to line location

Robert Schieweck

20.11.12

Ein wichtiges Problem in der Statistik ist die Bestimmung einer Regressionsgeraden. Diese soll für Beobachtungen $(x_i, y_i) \in \mathbb{R}^2$, $i = 1, \dots, n$ einen linearen Zusammenhang $y_i \approx sx_i + b$ für $i = 1, \dots, n$ beschreiben. Ein weiteres Problem ist die Einteilung von gegebenen Beobachtungen $(x_i, y_i) \in \mathbb{R}^2$ in Klassen (clustering), innerhalb derer die Beobachtungen “ähnlich” sind. Kombiniert man beide Probleme, erhält man ein Regressionproblem mit latenten Klassen: Die Beobachtungen folgen nicht einem einzelnen linearen Zusammenhang, sondern zerfallen in eine Anzahl K von Klassen, innerhalb derer jeweils ein eigener linearer Zusammenhang besteht. Es werden geometrische Lösungsverfahren für dieses kombinierte Problem vorgestellt.