

Aufgabe 1: Gegeben sei eine Menge X , auf der eine Abstandsfunktion d definiert ist. x_1, \dots, x_n seien Punkte, die einigermaßen gleichmäßig in X verteilt sind. Man kann nun einen Punkt y in X dadurch charakterisieren, welches x_i ihm am nächsten, am zweitnächsten, am drittnächsten usw. liegt. Definieren Sie einen Kern K (im Sinne der SVM-Theorie), der zwei Punkte y und z als ähnlich klassifiziert, wenn sie sich im Sinne dieser Charakterisierung ähneln.

Aufgabe 2: Beweisen oder widerlegen Sie die Behauptungen, die in der Vorlesung über die Ridge-Regression aufgestellt wurden (siehe v9.pdf, Seiten 4 bis 7).

Aufgabe 3: Simulieren Sie Daten, wie sie aus einer Microarray-Versuchserie mit unterschiedlichen Gewebetypen kommen könnten. Demonstrieren Sie anhand dieser Daten die Unterschiede zwischen Ridge-Regression, Lasso, Gene-Shaving etc.

Aufgabe 4: Besorgen Sie sich einen echten Genexpressionsdatensatz und erproben Sie an diesem die Analyseprogramme aus dem Bioconductor-Projekt.

www.ncbi.nlm.nih.gov/projects/geo

www.bioconductor.org