

Die Aufgabenstellung dieser Übung bringt eine Vertiefung im Gebiet des Data-Minings, im Speziellen wird die Methode der Principal Components Analysis (PCA) detailliert behandelt. Im File ‚hefe.mat‘ sind die Auswertungen der Spots eines DNA-Arrays enthalten. Es wurden 7 Arrays mit verschiedenen Experimenten ausgewertet. Die Daten eines Chips (1000 Spots) sind spaltenweise angeordnet, d.h. jede Spalte enthält die Daten eines Experiments. Wir werden untersuchen, ob die 7 Experimente auf gemeinsame Faktoren zurückgeführt werden können.

1. Laden Sie die Daten mit dem Befehle *load hefe.mat*.
 2. Erzeugen Sie einen Scatterplot des Experimente: (1,2), (1,3), (1,3).
Untersuchen Sie auch die Standardabweichungen der einzelnen Experimente. Können Sie aus den Scatterplots oder Standardabweichungen Unterschiede oder Gemeinsamkeiten der Experimente erkennen? [plot(), std()]
 3. Führen Sie die PCA entsprechend der Anleitung in der Vorlesung aus und wählen Sie die drei höchsten PCs aus. Die Eigenwerte sind ansteigend sortiert.
Hinweis: Verwenden Sie zur Berechnung der Kovarianzmatrix die Funktion *cov()*. Die Eigenwerte und Vektoren werden mittels *eig()* berechnet. Die Eigenwerte sind entlang einer Diagonalmatrix angeordnet.
 4. Berechnen Sie die z-Scores bezüglich dieser drei PCs für alle Spots und stellen Sie diese Scores wie in Aufgabe 2 dar. Untersuchen Sie auch die Standardabweichungen.
-