## 1. Kovarianzmatrix

### **a)** Code: ourCov.m

### b)

Im Datensatz 1 ist die Varianz in x-Richtung mit 81.4 recht hoch, in y-Richtung mit 1.18 vergleichsweise gering. Die Kovarianz liegt bei 1.78, somit besteht ein monotoner Zusammenhang zwischen den Daten. Außerdem sind die Daten annähernd achsenparallel.

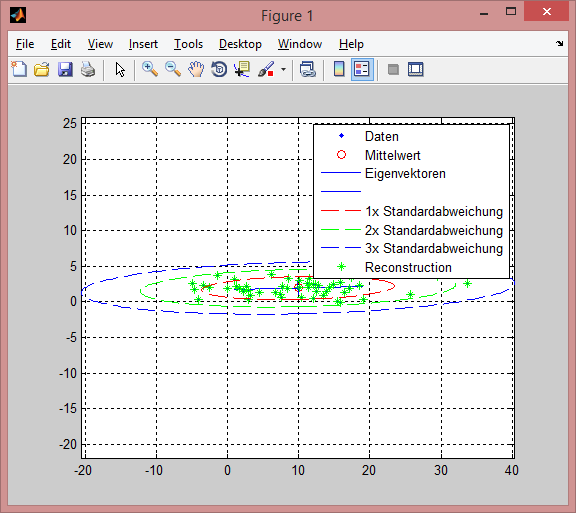
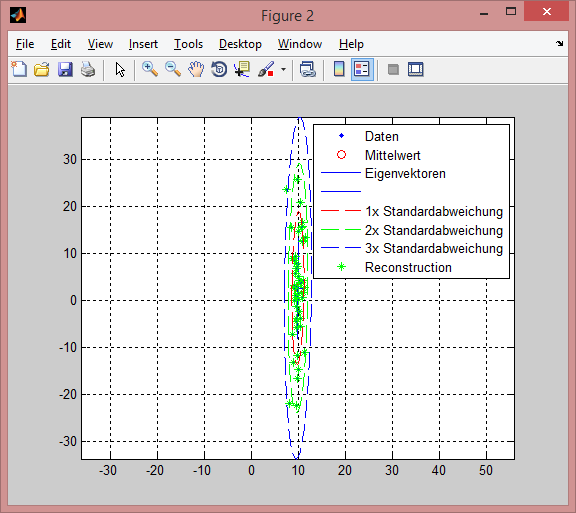
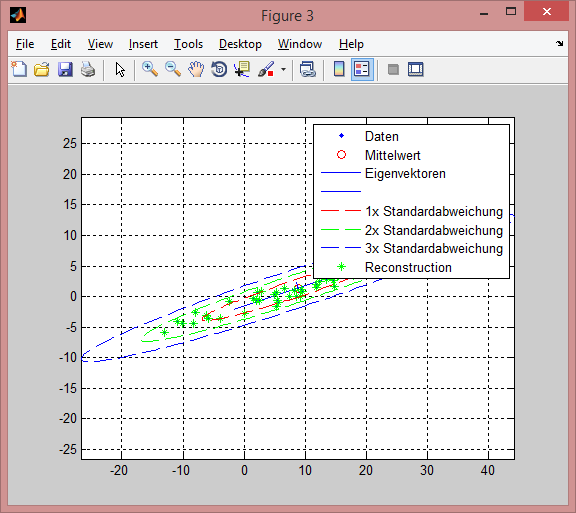
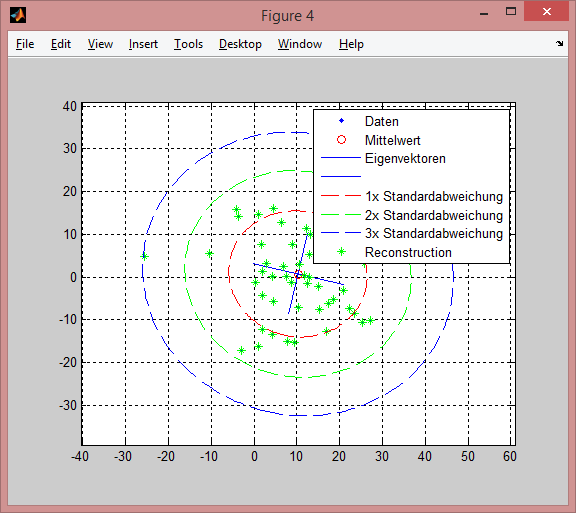
Im Datensatz 2 ist die Varianz in x-Richtung mit 0.74 sehr niedrig, in y-Richtung 116.31 vergleichsweise hoch. Die Kovarianz liegt bei 1.2, somit besteht ein monotoner Zusammenhang zwischen den Daten. Außerdem sind die Daten annähernd achsenparallel.

Im Datensatz 3 ist die Varianz in x-Richtung mit 110.49 recht hoch, in y-Richtung 12.75 vergleichsweise gering. Die Kovarianz liegt bei 36.1, somit besteht ein monotoner Zusammenhang zwischen den Daten, und sie sind auf einer monoton steigenden Linie angeordnet.

Im Datensatz 4 liegt die Varianz in x-Richtung bei 115.84, in y-Richtung bei 96.58, die Streuung der Daten ist in beide Richtungen sehr hoch. Die Kovarianz liegt bei -4.5, somit besteht kein monotoner Zusammenhang zwischen den Daten.

## 2. PCA

### **a)**

### **b)**

Eigenvektor: Vektor, dessen Richtung in der Abbildung nicht verändert wird, sondern nur skaliert.

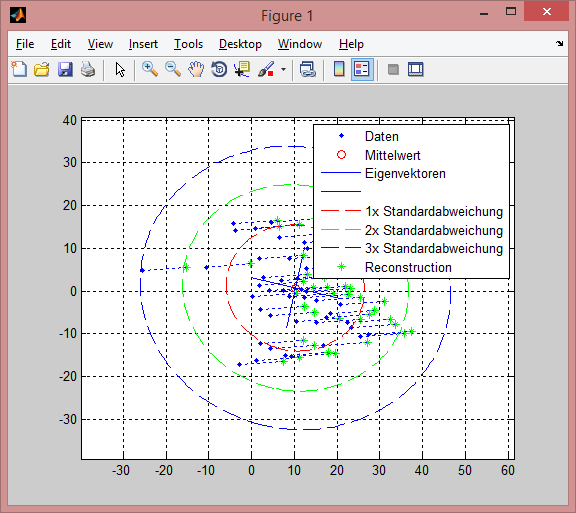
Die Eigenvektoren entsprechen der Kovarianz und geben die Lage der statistischen Verteilung im Raum/der Ebene an.

### **c)**

Der Eigenwert gibt den Skalierungsfaktor des Eigenvektors an. Er entspricht der Varianz und gibt die Skalierung der Verteilung in Richtung der Eigenvektoren an**.**

### **d)**

Das Nichtabziehen des Mittelwerts hätte zur Folge, dass die Datenpunkte um die Mittelwerts-Vektoren verschoben wären und stimmen somit nicht mehr dem ursprünglichen Datensatz überein.



## 3. Unterraum-Projektion

### a)

(ich versteh nicht ganz, was die Fragestellung von mir will)

Im Grunde ist das jetzt eindimensional, weil alles auf einer Linie liegt.

Der durchschnittliche Fehler liegt bei 0.7257259304%

### b)

Die Projektion versteh ich nicht so ganz

Nutzt man den Nebenvektor anstelle des Hauptvektors, liegt der durchschnittliche Fehler bei 8.9%. Es empfiehlt sich also, jene Eigenvektoren zu verwenden, die die größte Varianz besitzen, da beim Projizieren auf einen Vektor mit geringer Varianz der Fehler steigt.

## 4. Untersuchungen in 3D

### a)

Relation? :/

### b)

Projektion?

Werden die Daten auf den Unterraum der ersten beiden Eigenvektoren (= jene mit der höchsten Ausdehnung) projiziert, so liegen die Daten auf einer Ebene statt im dreidimensionalen Raum und es geht eine Ausdehnung der Daten verloren.

## 5. Shape Modell

### a) Code: Bsp5.m

### b)

1. Mode in Range +-3\*105.4005: Unterschiede zum mean shape relativ groß, ähnliche Form aber erkennbar.

4. Mode in Range +-3\*13.0136: Gute Annäherung im linken Bereich des Shapes, große Abweichungen auf der rechten Seite, vor allem in den Kurven

5. Mode in Range +-3\*10.8447: Gute Annäherung im rechten Bereich des Shapes, große Abweichungen im linken Bereich, vor allem auch eher geraden Strecken

7. Mode in Range +-3\*5.7134: Gute Annäherung im oberen Bereich des Shapes, Abweichungen im unteren Bereich, vor allem in senkrechten Bereichen

12. Mode in Range +-3\*2.5475: Shapes entsprechen in Größe und Form weitestgehend dem mean shape, nur noch geringe Unterschiede feststellbar

### c)

Haupteigenvektor: entspricht 80% der Varianz

Nebeneigenvektor: zahlenabhängig kann das Ergebnis eine Annäherung an das Mean Shape sein, in den meisten Fällen ist die Form kaum erkennbar

Ersten 2 Eigenvektoren: entspricht 90% der Varianz

100% der Varianz = 14 Eigenvektoren: Wirres Gekrakel, nicht verwertbar

95% der Varianz = 3 Eigenvektoren: Unter Umständen wirres Gekrakel

90% der Varianz = 2 Eigenvektoren: zahlenabhängig kann das Ergebnis eine Annäherung an das Mean Shape sein, in den meisten Fällen ist die Form kaum erkennbar

80% der Varianz = 1 Eigenvektor: relativ gute Annäherung