Maschinelles Lernen Blatt 5

Nikolas Zeitler, Joshua Hartmann, Alexander Diegel

June 8, 2016

1 Fragen zur Vorlesung

a) Warum wendet man die Principal Component Analysis (PCA) an? (Zwei Gründe sollten genannt werden.)

1. Dimensionsreduktion komplexer/multidimensionaler/hochdimensionaler Datensätze(durch Linearkombination von Merkmalen), falls Verarbeitung aller vorhandener Dimensionen zu komplex wäre.

2. Mit der PCA kann man die Dimensionen des Datensatzes finden, die die meisten Informationen über die Daten enthalten. (Rauschen wird dabei ignoriert.)

b) Was ist der Unterschied zwischen der PCA und der (Fisher) Linear Discriminant Analysis (LDA)?

PCA wird zur Repräsentation von Daten genutzt genutzt.
LDA hingegen wird zur Klassifikation genutzt.

c) Welcher Abstand soll von der LDA maximiert werden?

An sich soll der Abstand der Mittelwerte (im Verhältnis zur Streuung der Klassen) maximiert werden.

d) Warum genügt es nicht, den eben genannten Abstand zu maximieren?

Es ist auch wichtig, wie weit gestreut wird.

Es ist z.B. oft besser, wenn man sehr scharfe Verteilungen erhält deren Mittelwert näher beieinander liegen als Verteilungen, deren Mittelwerte zwar weiter auseinander liegen, die aber wegen großer Streuung doch stark überlappen.

e) Was beschreibt die within-class scatter matrix (Foliensatz 5, Folie 17)? Die within-class scatter matrix beschreibt die Streuung/Varianz innerhalb einer Klasse (wir wollen, dass die Punkte einer Klasse möglichst eng beieinander liegen).

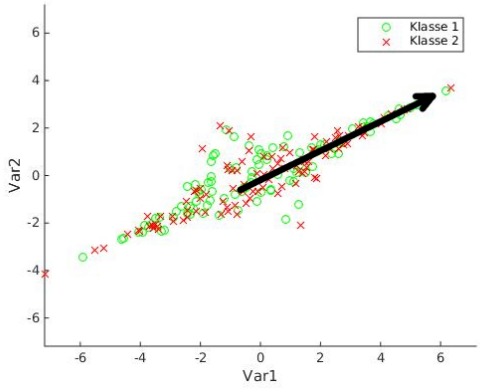
f) Was beschreibt die between-class scatter matrix (Foliensatz 5, Folie 18)? Die between-class scatter matrix beschreibt den Abstand der projezierten Klassen-Mittelwerte (dieser Abstand sollte möglichst groß sein).

g) Man gebe sowohl für PCA als auch LDA an, ob es sich um ein supervised oder unsupervised Lernverfahren handelt.

PCA: Unsupervised
LCA: Supervised

2 Principal Component Analysis – 1

(a)



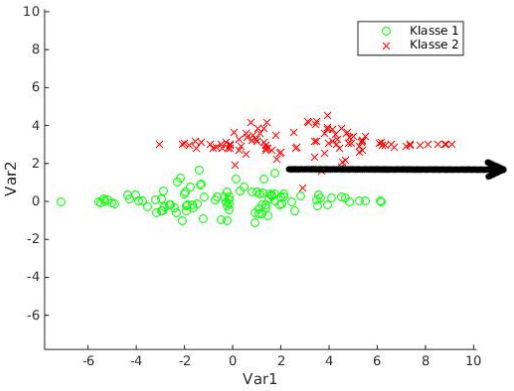
1. Bietet sich im Allgemeinen eine Dimensionsreduktion auf den PC-Vektor an?

   Die Daten befinden sich relativ gut auf einer Geraden, also bietet sich eine Dimensionsreduktion an.

2. Ist dies für die Klassifikation förderlich?

   Nein ist es nicht, durch eine Reduktion sind die Klassen nicht gut unterscheidbar.

(b)



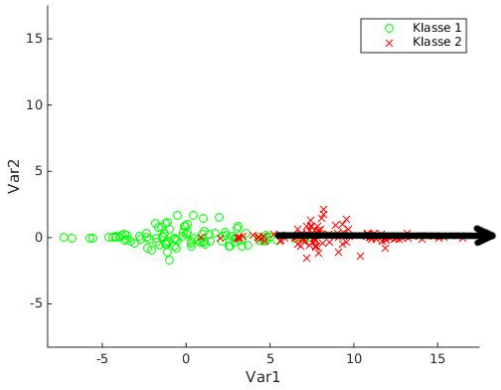
1. Bietet sich im Allgemeinen eine Dimensionsreduktion auf den PC-Vektor an?

   Die Daten befinden sich tendenziell eher auf einer Linie. Eine Dimensionsreduktion könnte sich durchaus anbieten, falls der Verlust des Var2 Wertes hinnehmbar ist, denn die Daten unterscheiden sich stark im Bezug auf den Var1 Wert.

2. Ist dies für die Klassifikation förderlich?

   Nein ist es nicht, durch eine Reduktion sind die Klassen nicht gut unterscheidbar.

(c)



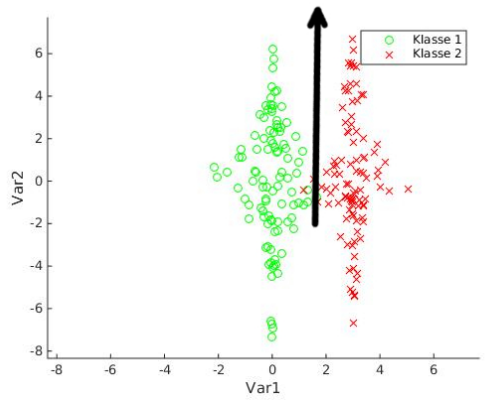
1. Bietet sich im Allgemeinen eine Dimensionsreduktion auf den PC-Vektor an?

   Ja, die Daten unterscheiden sich hauptsächlich im Wert von Var1, eine Dimensionsreduktion bietet sich also an.

2. Ist dies für die Klassifikation förderlich?

   Ja ist es, durch eine Reduktion sind die Klassen immer noch gut unterscheidbar.

(d)



1. Bietet sich im Allgemeinen eine Dimensionsreduktion auf den PC-Vektor an?

   Die Daten befinden sich tendenziell eher auf einer Linie. Eine Dimensionsreduktion könnte sich durchaus anbieten, falls der Verlust des Var1 Wertes hinnehmbar ist, denn die Daten unterscheiden sich stark im Bezug auf den Var2 Wert.

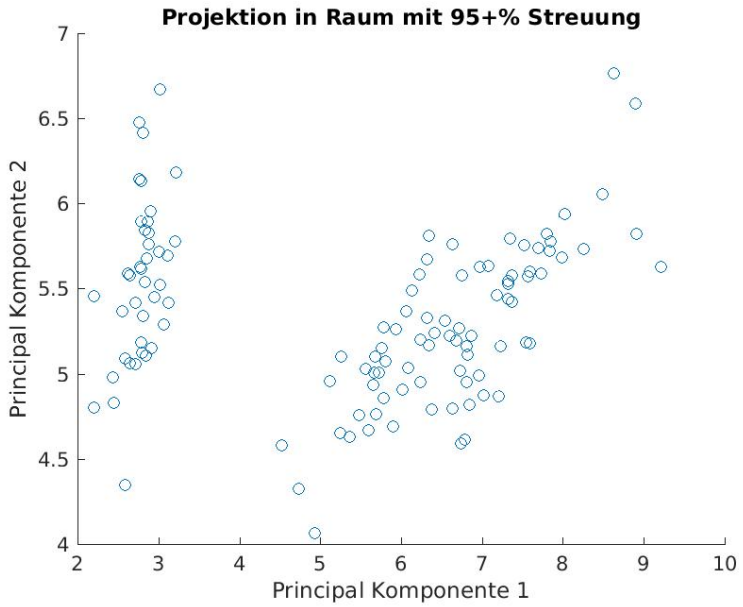
2. Ist dies für die Klassifikation förderlich?

   Nein ist es nicht, durch eine Reduktion sind die Klassen nicht gut unterscheidbar.

3 Principal Component Analysis – 2

Welche PCs enthalten zusammen mindestens 95% der Streuung?
Komponente 1 und 2 enthalten 97.6% der Streuung (die erste Komponente allein macht ca. 92% aus). Die erste und zweite Komponente werden auf jeden Fall benötigt um die 95% zu erreichen. Komponente 3 und 4 haben fast keinen Einfluss auf die Streuung.

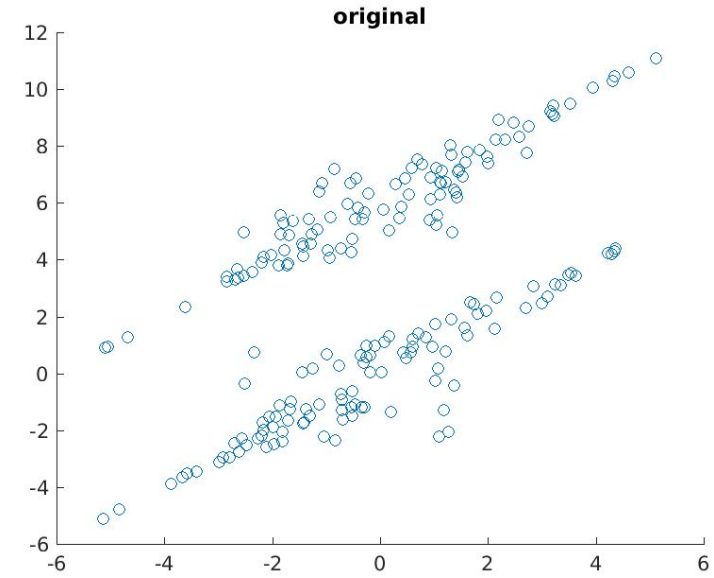
Plotten des Ergebnisses

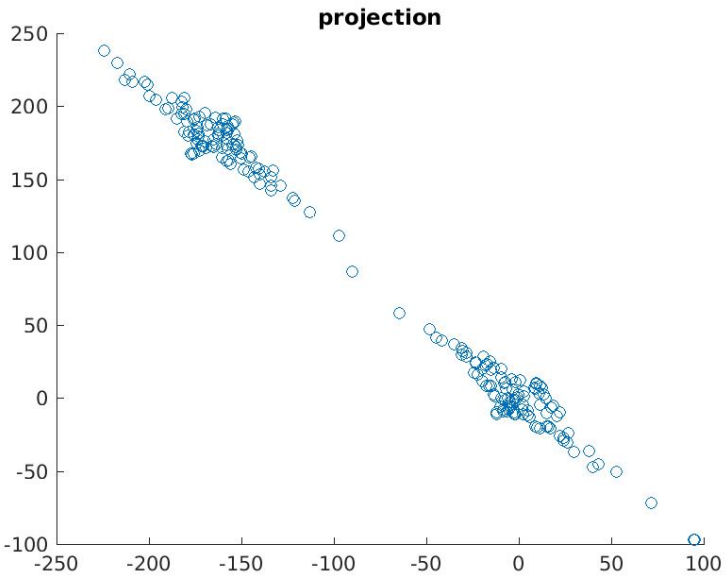


Ergibt die PCA in diesem Zusammenhang Sinn? Warum oder warum nicht?
Ja - wir sehen, dass es möglich ist unseren Datensatz durch eine Projektion in die entsprechende Richtungen (erste beiden Principal Components) zu separieren, in diesen beiden Richtungen/Komponenten ist die meiste Information des Datensatzes enthalten. Fuer eine Klassifikation bzw. Zuordnung der Pflanzen anhand der Merkmale macht diese Projektion aber keinen Sinn mehr.

4 Linear Discriminant Analysis

d) Plottet die originalen Datensätze und deren Projektionen und beurteilt bzw. begründet die Resultate.





Ziel der LDA ist es die Klassen so zu separieren, dass deren jeweiligen Mittelwerte weit auseinander liegen. Gleichzeitig soll die Varianz aber möglichst klein sein, dh. Punkte einer Klasse sollen eng beieinander auf der Projektion liegen. Dies ist gut zu erkennen. Die obere Klasse in der Projektion entspricht der oberen Klasse der orginal Daten. Die

Zentren beider Klassen sind in der Projektion weiter auseinander als im Orginal und die Varianz relativ gesehen kleiner.