

1. Aufgabe

Wir untersuchen 367 Bodenproben mittels einer chemischen Analyse, welche jede der 367 Proben auf 17 Chemikalien testet. Alle Chemikalien werden in derselben Einheit und auf derselben Skala gemessen.

Die Daten befinden sich im data frame `dat` im R-data file `ueb102001.rda`.

Zur weiteren Bearbeitung der Daten möchten wir die Dimensionalität so weit wie möglich vereinfachen. Wir möchten danach aber immer noch einen grossen Anteil der Variabilität der Daten erklären.

Benutzen Sie PCA um die Dimensionalität dieses Datensatzes zu reduzieren. Welche der folgenden Aussagen sind korrekt?

- (a) Die 5. PC erklärt etwa 42.1 Prozent der totalen Varianz.
- (b) Die Mindestanzahl an PCs, die man benötigt, um mehr als 60 Prozent der totalen Variabilität zu erklären ist 3.
- (c) Die Koordinate der 8. PC vom Datenpunkt 99 lautet -0.2. (Hinweis: Beachten Sie nur den Wert, das Vorzeichen ist willkürlich)

Lösung

```
> # a)
> pr.out <- prcomp(dat)
> myVar <- pr.out$sdev^2

> myVar[5] / sum(myVar)
> 2

> # b)
> rr <- cumsum(myVar) / sum(myVar)

> which(rr > 0.6)[1]
> 2

> # c)

> pr.out$x[99,8]
> -0.2
```

- (a) **False.** Die 5. PC erklärt etwa 2 Prozent der totalen Varianz.
- (b) **False.** Die Mindestanzahl an PCs die man benötigt ist 2.
- (c) **True.** Die Koordinate der 8. PC vom Datenpunkt 99 lautet -0.2.

2. Aufgabe

Wir untersuchen 450 Bodenproben mittels einer chemischen Analyse, welche jede der 450 Proben auf 14 Chemikalien testet. Alle Chemikalien werden in derselben Einheit und auf derselben Skala gemessen.

Die Daten befinden sich im data frame `dat` im R-data file `ueb603147.rda`.

Zur weiteren Bearbeitung der Daten möchten wir die Dimensionalität so weit wie möglich vereinfachen. Wir möchten danach aber immer noch einen grossen Anteil der Variabilität der Daten erklären.

Benutzen Sie PCA um die Dimensionalität dieses Datensatzes zu reduzieren. Welche der folgenden Aussagen sind korrekt?

- (a) Die 6. PC erklärt etwa 2 Prozent der totalen Varianz.
- (b) Die Mindestanzahl an PCs, die man benötigt, um mehr als 90 Prozent der totalen Variabilität zu erklären ist 9.
- (c) Die Koordinate der 2. PC vom Datenpunkt 355 lautet 14.78. (Hinweis: Beachten Sie nur den Wert, das Vorzeichen ist willkürlich)

Lösung

```
> # a)
> pr.out <- prcomp(dat)
> myVar <- pr.out$sdev^2

> myVar[6] / sum(myVar)
> 2
```

```
> # b)
> rr <- cumsum(myVar) / sum(myVar)

> which(rr > 0.9)[1]
> 7
```

```
> # c)

> pr.out$x[355,2]
> -13.49
```

- (a) **True**. Die 6. PC erklärt etwa 2 Prozent der totalen Varianz.
- (b) **False**. Die Mindestanzahl an PCs die man benötigt ist 7.
- (c) **False**. Die Koordinate der 2. PC vom Datenpunkt 355 lautet -13.49.

3. Aufgabe

Wir untersuchen 491 Bodenproben mittels einer chemischen Analyse, welche jede der 491 Proben auf 16 Chemikalien testet. Alle Chemikalien werden in derselben Einheit und auf derselben Skala gemessen.

Die Daten befinden sich im data frame `dat` im R-data file `ueb23847.rda`.

Zur weiteren Bearbeitung der Daten möchten wir die Dimensionalität so weit wie möglich vereinfachen. Wir möchten danach aber immer noch einen grossen Anteil der Variabilität der Daten erklären.

Benutzen Sie PCA um die Dimensionalität dieses Datensatzes zu reduzieren. Welche der folgenden Aussagen sind korrekt?

- (a) Die 1. PC erklärt etwa 3.7 Prozent der totalen Varianz.
- (b) Die Mindestanzahl an PCs, die man benötigt, um mehr als 80 Prozent der totalen Variabilität zu erklären ist 5.

- (c) Die Koordinate der 1. PC vom Datenpunkt 322 lautet -20.26. (Hinweis: Beachten Sie nur den Wert, das Vorzeichen ist willkürlich)

Lösung

```
> # a)
> pr.out <- prcomp(dat)
> myVar <- pr.out$sdev^2
> myVar[1] / sum(myVar)
> 20.5
```

```
> # b)
> rr <- cumsum(myVar) / sum(myVar)
> which(rr > 0.8)[1]
> 5
```

```
> # c)
> pr.out$x[322,1]
> -13.68
```

- (a) **False.** Die 1. PC erklärt etwa 20.5 Prozent der totalen Varianz.
(b) **True.** Die Mindestanzahl an PCs die man benötigt ist 5.
(c) **False.** Die Koordinate der 1. PC vom Datenpunkt 322 lautet -13.68.