GRUNDLAGEN DES MASCHINELLEN LERNENS

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

Marvin Häußermann (`marvin.haeussermann@uni-tuebingen.de`)
Tobias Lang (`t.lang@uni-tuebingen.de`)
Mathias Schickel (`mathias.schickel@uni-tuebingen.de`)

Andreas Schilling
Sommersemester 2020

Blatt 4

Ausgabe: 12.06.2020; Abgabe: bis 26.06.2020, 09:00 Uhr.

Wer die Geometrie begreift, vermag in dieser Welt alles zu verstehen.
GALILEO GALILEI

Organisatorisches und Formales

- Es gibt *genau ein* Gruppenmitglied der *Vierer*gruppe die Lösung in Form eines Archivs ab, am besten im `zip`-Format.
- Das Archiv und der darin enthaltene Ordner mit allen Dateien, die zur Abgabe gehören, sowie das dort mitgelieferte PDF(!) mit der Textabgabe müssen der Namenskonvention `nachname1-nachname2-nachname3-nachname4-uebungX` entsprechen. Sonderzeichen sollten dabei nicht verwendet werden und die Namen sollen alphabetisch sortiert sein, z. B. `maier-mueller-schmidt-schulz-uebung2.zip`. (Die Zahl der angegebenen Namen soll dabei der Größe der Gruppe entsprechen.)
- Auf gute und prägnante Formulierung sowie Rechtschreibung und Zeichensetzung ist zu achten. Ausschweifende Antworten sowie mangelnde sprachliche Sorgfalt können mit Punktabzug geahndet werden.
- Textabgaben im Code werden **nicht gewertet**.
- Abbildungen dürfen **nicht übereinander geplottet** werden, außer es wird explizit verlangt.
- Insofern Dateien mit Funktionsrümpfen zur Verfügung gestellt wurden, sind diese beim Verfassen des Programmcodes zu verwenden und **in der vorgegebenen Ordnerstruktur** wieder abzugeben. Die Form der Ein- und Ausgabeparameter darf dabei **nicht verändert werden**.

Aufgabe 1 (Hauptkomponentenanalyse und lineare Diskriminanzanalyse) (15 Punkte)

a) Bei der *Hauptkomponentenanalyse* (Englisch *Principal Component Analysis*, kurz *PCA*) werden die Daten neu repräsentiert. Was sind dabei die zwei wesentlichen Ziele? Respektiert die Repräsentation im Allgemeinen Klassenzugehörigkeiten? (3 Punkte)

b) Was ist das Ziel der *linearen Diskriminanzanalyse* (Englisch *(Fisher) Linear Discriminant Analysis*, kurz *LDA*)? Erkläre dabei kurz, was die Gemeinsamkeit mit der PCA ist und worin der Unterschied besteht. (2 Punkte)

c) Angenommen, es liegen d-dimensionale Daten von c Klassen vor ($d > c$). Welche Dimension haben die Daten in der Repräsentation durch die LDA? (1 Punkt)

d) Bei der LDA wird das Verhältnis zweier Größen maximiert. Um welche Größen handelt es sich und weswegen würde es nicht genügen, nur eine der beiden Größen zu betrachten? (Überlege dazu, was das Ziel der LDA ist.) (2 Punkte)

e) Welche der beiden in der letzten Teilaufgabe genannten Größen kann mittels der *Interklassenstreumatrix* S_B (*between-class scatter matrix*) gemessen werden und welche mittels der *Innerklassenstreumatrix* S_W (*within-class scatter matrix*)?

   Über welche Formeln sind diese Matrizen jeweils definiert? (Es darf der Fall, dass nur zwei Klassen vorliegen, gewählt werden.)

   Hängen die Matrizen selbst von dem Unterraum ab (und damit von den diesen Raum aufspannenden Vektoren), auf den man die Daten im Zuge der LDA projizieren möchte? (3 Punkte)

f) Gib für den Fall zweier Klassen das Funktional $J(\boldsymbol{w})$ in Termen der Matrizen S_B und S_W an, das für die LDA über den Vektor $\boldsymbol{w}$ maximiert werden soll. Welche Rolle spielt der Vektor $\boldsymbol{w}$ für die LDA? (2 Punkte)

g) Man gebe sowohl für die PCA als auch die LDA an, ob es sich um ein *supervised* oder *unsupervised* Lernverfahren handelt, und begründe die Antwort. (2 Punkte)

Aufgabe 2 (Hauptkomponentenanalyse – 1) (15 Punkte)

Bei der Hauptkomponentenanalyse werden die Datenpunkte $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \mathbb{R}^m$ auf einen affinen Unterraum der Dimension $d < m$ projiziert.

Weise für den Fall $d = 1$ nach, dass die Richtung $\boldsymbol{e}$ der Geraden durch den Datenmittelwert $\boldsymbol{m}$ durch einen zum größten Eigenwert der *Streumatrix* S gehörenden Eigenvektor gegeben ist. Dazu soll gezeigt werden, dass mit dieser Wahl von $\boldsymbol{e}$ mit $\|\boldsymbol{e}\| = 1$ das Funktional

$$J_1(\boldsymbol{e}) := \sum_{k=1}^{n} \left\|(\boldsymbol{x}_k - \boldsymbol{m} \,|\, \boldsymbol{e}) \cdot \boldsymbol{e} - (\boldsymbol{x}_k - \boldsymbol{m})\right\|^2$$

minimiert wird.

Die Streumatrix S ist durch $\sum_{k=1}^{n} (\boldsymbol{x}_k - \boldsymbol{m})(\boldsymbol{x}_k - \boldsymbol{m})^\intercal$ gegeben. Es darf ohne Beweis verwendet werden, dass für eine symmetrische reelle Matrix S der Ausdruck $\boldsymbol{x}^\intercal S \boldsymbol{x}$ maximal ist, wenn $\boldsymbol{x}$ ein Eigenvektor zum größten Eigenwert der Matrix S ist.

Hinweise

1. Zunächst sollte das Funktional unter Beachtung der nachfolgenden Tipps zum Skalarprodukt geeignet umgeformt werden, sodass es durch die Streumatrix (und einen weiteren von $\boldsymbol{e}$ unabhängigen Term) ausgedrückt werden kann. Danach ist der obige Hinweis zu symmetrischen, reellen Matrizen zu beachten, um den Beweis zu vervollständigen.

2. $(\boldsymbol{x} \,|\, \boldsymbol{y}) = \boldsymbol{x}^\intercal \boldsymbol{y} \in \mathbb{R}$ bezeichnet das bilineare und symmetrische Standardskalarprodukt, für das $\|\boldsymbol{x}\|^2 = (\boldsymbol{x} \,|\, \boldsymbol{x})$, $(\alpha\boldsymbol{x} + \boldsymbol{y} \,|\, \boldsymbol{z}) = \alpha(\boldsymbol{x} \,|\, \boldsymbol{z}) + (\boldsymbol{y} \,|\, \boldsymbol{z})$ sowie $(\boldsymbol{x} \,|\, \boldsymbol{y}) = (\boldsymbol{y} \,|\, \boldsymbol{x})$ gilt.

3. Es gilt mit der Bilinearität $(\boldsymbol{x} - \boldsymbol{y} \,|\, \boldsymbol{x} - \boldsymbol{y}) = (\boldsymbol{x} \,|\, \boldsymbol{x}) - 2(\boldsymbol{x} \,|\, \boldsymbol{y}) + (\boldsymbol{y} \,|\, \boldsymbol{y})$.

4. Es gilt $(\boldsymbol{x} \,|\, \boldsymbol{y})^2 = (\boldsymbol{x}^\intercal \boldsymbol{y})(\boldsymbol{y}^\intercal \boldsymbol{x}) = \boldsymbol{x}^\intercal (\boldsymbol{y}\boldsymbol{y}^\intercal)\boldsymbol{x}$, wobei $\boldsymbol{y}\boldsymbol{y}^\intercal$ das äußere Produkt des Vektors $\boldsymbol{y}$ mit sich selbst ist (und damit eine symmetrische reelle Matrix).

5. Es ist sinnvoll, $\|\dots\|$ im Funktional in $(\dots \,|\, \dots)$ umzuformen.

6. Man sollte zunächst auf die Formel

   $$\sum_{k=1}^{n} \Big[\big((\boldsymbol{x}_k - \boldsymbol{m} \,|\, \boldsymbol{e}) \cdot \boldsymbol{e} \,|\, (\boldsymbol{x}_k - \boldsymbol{m} \,|\, \boldsymbol{e}) \cdot \boldsymbol{e}\big) - 2\big((\boldsymbol{x}_k - \boldsymbol{m} \,|\, \boldsymbol{e}) \cdot \boldsymbol{e} \,|\, \boldsymbol{x}_k - \boldsymbol{m}\big) + \big(-(\boldsymbol{x}_k - \boldsymbol{m}) \,|\, -(\boldsymbol{x}_k - \boldsymbol{m})\big)\Big]$$

   hinarbeiten und von da aus weiter umformen. Zum Beispiel gilt (mit $(\alpha\boldsymbol{x} \,|\, \boldsymbol{y}) = \alpha(\boldsymbol{x} \,|\, \boldsymbol{y})$)

   $$2\big((\boldsymbol{x}_k - \boldsymbol{m} \,|\, \boldsymbol{e}) \cdot \boldsymbol{e} \,|\, \boldsymbol{x}_k - \boldsymbol{m}\big) = 2(\boldsymbol{e} \,|\, \boldsymbol{x}_k - \boldsymbol{m})^2$$

   und

   $$\big(-(\boldsymbol{x}_k - \boldsymbol{m}) \,|\, -(\boldsymbol{x}_k - \boldsymbol{m})\big) = \|\boldsymbol{x}_k - \boldsymbol{m}\|^2.$$

7. Man sollte $(\boldsymbol{x} \,|\, \boldsymbol{y})$ erst möglichst spät durch $\boldsymbol{x}^\intercal \boldsymbol{y}$ ersetzen und vorher stets mit den angegebenen Regeln zum Skalarprodukt umformen.

Aufgabe 3 (Hauptkomponentenanalyse – 2) (15 Punkte)

1. Betrachte die folgenden Grafiken zu Merkmalen jeweils zweier Klassen und schätze den Mittelwert der Daten und die Richtung des vom Mittelwert ausgehenden ersten Hauptkomponentenvektors (für *alle* Datenpunkte ohne Klassenunterschied) und zeichne ihn ein. Erkläre,
   a) ob sich im Allgemeinen (d. h. unabhängig von der Klasse) eine Dimensionsreduktion auf den durch diese Richtung gegebenen affinen Unterraum anbietet und
   b) ob die resultierende Repräsentation der Daten die Trennbarkeit der Klassen verschlechtert. (12 Punkte)

2. Wähle eine der Grafiken aus, bei denen die PCA im Hinblick auf die Klassifizierung der projizierten Datenpunkte nicht sinnvoll ist. Erkläre und veranschauliche anhand dieser Grafik den Unterschied zur LDA, indem Du zusätzlich zum bereits eingezeichneten Hauptkomponentenvektor der PCA den Vektor einzeichnest, auf den die Daten bei der LDA projiziert werden. (3 Punkte)

Die folgenden Plots werden auch als PDF-Dateien mit dem Übungsblatt zur Verfügung gestellt.

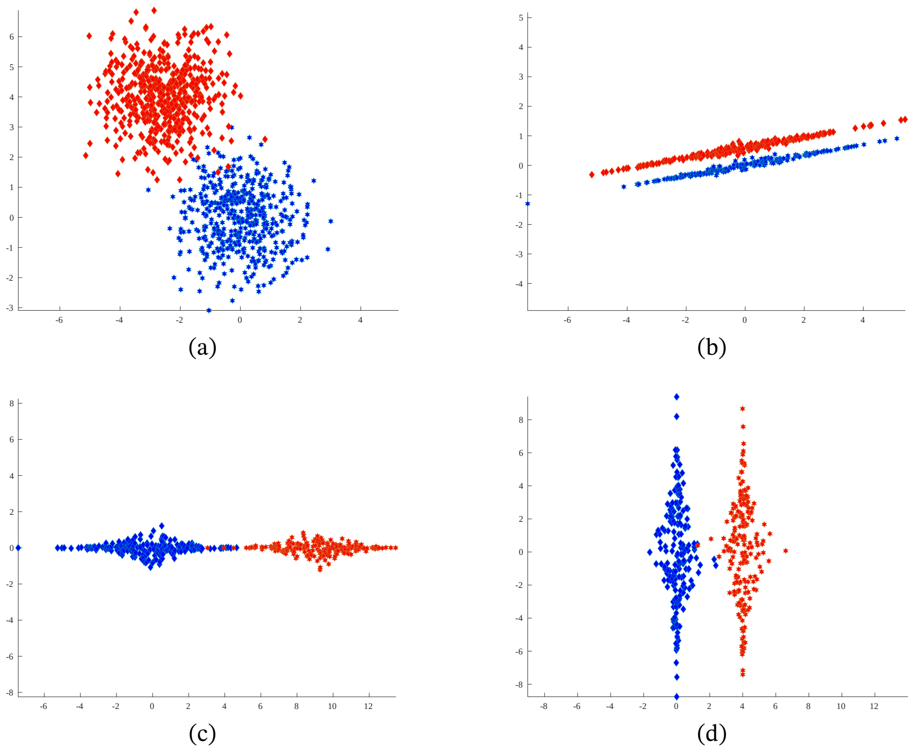

Abbildung 1: Plots

Aufgabe 4 (Hauptkomponentenanalyse – 3) (10 Punkte)

In dieser Aufgaben wird nochmals mit dem aus Übungsblatt 2 bekannten Schwertlilien-Datensatz gearbeitet. Diesmal soll allerdings nicht klassifiziert, sondern untersucht werden, welche Hauptkomponenten den meisten Informationsgehalt liefern. Dazu sollen die folgenden Schritte befolgt werden:

a) Lese die Trainingsdaten ein und ignoriere dabei die Klassen und behandle alle Datensätze als einen einzigen.

b) Führe eine PCA durch. (2 Punkte)

c) Welche Hauptkomponenten enthalten zusammen mindestens 95 Prozent der Streuung? (2 Punkte)

d) Projiziere die vierdimensionalen Daten in einen Raum, der mindestens 95 Prozent der Streuung enthält. (2 Punkte)

e) Plotte das Ergebnis geeignet. Die verschiedenen Klassen sollen dabei unterschiedlich dargestellt werden. (2 Punkte)

f) Ist es im vorliegenden Fall sinnvoll, die Hauptkomponentenanalyse anzuwenden? Erkläre die gegebene Antwort. (Es sind unterschiedliche Auffassungen denkbar.) (2 Punkte)

Hinweise In Matlab kann für eine `n*m` Matrix `data` mit `n` Zeilen und `m` Spalten über den Befehl `[pc ws scatter] = pca( data )` die PCA auf die Daten in der Matrix angewandt werden, wobei die Spalten Kenngrößen repräsentieren und die Zeilen Stichproben. Dabei liefert `pc` eine Matrix mit Spaltenvektoren, die die Hauptkomponenten (*principal components*) repräsentieren. In `scatter` wird für jede Hauptkomponente angegeben, wieviel Streuung der Datensatz auf sie projiziert aufweist. Die Variable `ws` („working set“) liefert die Koordinaten der Punktmengen im Raum der Projektion.

Man beachte, dass bei der PCA die Punktmenge im Ursprung zentriert wird. `ws` ist also *eine* Möglichkeit, die Projektion vorzunehmen. Es ist aber zu empfehlen, *nicht* mit der Variablen `ws` zu arbeiten: Wenn neue (Test-)Daten verwendet werden sollen, müssen diese sonst ebenfalls erst im Ursprung zentriert werden und danach reduziert werden. Eine andere Möglichkeit besteht demgegenüber darin, den ursprünglichen Datensatz auf die gewünschten PCs zu projizieren. Dann kann die Zentrierung um den Ursprung vermieden werden. Gearbeitet werden kann mit diesem Beispiel, das in Pseudocode angegeben ist:

```
data = [1 2; 3 4; 5 6];          // create some data
[pc ws sc] = pca( data );        // apply PCA
firstComp = pc( :, 1 );          // get first PC.
newData = firstComp' * data';    // project data onto first PC
newData = newData';              // rearrange to column order.
```

Aufgabe 5 (Lineare Diskriminanzanalyse) (10 Punkte)

In dieser Aufgabe soll eine *Diskriminanzanalyse* (LDA) durchgeführt werden. Mit dem Übungsblatt wird die Funktion

```
createPointCloud( numPoints, variance, stretchFactor, alpha )
```

zur Verfügung gestellt. Sie erstellt eine Punktmenge, die grundsätzlich entlang der x-Achse gestreckt ist. Die einzelnen Argumente sollen dabei hier kurz erläutert werden:

- `numPoints` bezeichnet die Anzahl der zu erstellenden Punkte,
- `variance` bezeichnet die Streuung in x-Richtung,
- `stretchFactor` beeinflusst die Streuung in y-Richtung,
- `alpha` gibt den Winkel (in Grad) an, um den die Punktwolke gedreht wird.

Folgende Aufgaben sind nun zu bearbeiten:

a) Es sollen zwei Punktmengen erstellt werden, wobei eine Separierbarkeit zwischen den beiden repräsentierten Klassen sichergestellt sei. Im Live-Skript `aufgabe5.mlx` ist dies bereits umgesetzt. Zur *eigenen Orientierung* sind Variationen aber erlaubt.

b) Führe eine LDA durch. (6 Punkte)

c) Projiziere die Punktmengen in den optimalen Raum. (2 Punkte)

d) Plotte die originalen Datensätze und deren Projektionen in ein gemeinsames Diagramm und beurteile bzw. begründe die Resultate. (2 Punkte)

Hinweise

1. Achte bei der Darstellung der Punktwolken darauf, dass die Achsenskalierung in x- und y-Richtung identisch ist.
2. Für die Lösung von Aufgabenteil b) wird folgendes Vorgehen vorgeschlagen:
   - Berechne μ_1, μ_2 und μ.
   - Berechne die *scatter-within* der Klassen 1 und 2 und die gesamte Streuung – in der Form `ScatterW = ScatterW1 + ScatterW2`.
   - Ermittle die *scatter-between* der Klassen.
   - Berechne die Projektionsmatrix `proj = inv( ScatterW ) * ScatterB` und beachte: Diese Projektion behält die Darstellung im zweidimensionalen Raum bei, alle Punkte befinden sich aber auf einer Linie.
   - Alternativ zum letzten Schritt kann die Matrix `proj = inv( ScatterW ) * (mu1 - mu2)'` bestimmt werden. Dann ist zu beachten, dass diese Projektion auf einen eindimensionalen Raum erfolgt.
3. Bei Matrix- bzw. Vektoroperationen ist es sehr wichtig, sich über die Ausrichtung der Daten im Klaren zu sein. In der Vorlesung wird von Spaltenvektoren ausgegangen, oftmals wird jedoch mit Zeilenvektoren gearbeitet.