

## Blatt 6

Abgabe bis Mittwoch, den 8. Dezember 2010 um 12:00

Die Lösungen bitte per Email an **irene.winkler@tu-berlin.de** senden  
 und im **Postfach** von Dr. Konrad Rieck abgeben.

## Lineare Fisher-Diskriminanzanalyse

Auf diesem Aufgabenblatt soll die Fisher-Diskriminanzanalyse (FDA) implementiert werden, und es soll visualisiert werden, wie die gefundene Lösung das zugrunde liegende Kriterium maximiert.

Zur Erinnerung: Wir haben einen Datensatz  $X_i, Y_i, 1 \leq i \leq n$ , mit  $Y_i = \pm 1$ . Sei  $D_k$  die Menge aller Punkte  $X_i$  so dass  $Y_i = k$ , und  $m_k = \sum_{x \in D_k} x / |D_k|$  der Mittelpunkt der Klasse  $k$ .

Die FDA-Kostenfunktion lautet

$$J(w) = \frac{w^\top S_B w}{w^\top S_W w},$$

wobei

$$S_W = S_1 + S_2 \quad \text{mit} \quad S_i = \sum_{x \in D_i} (x - m_i)(x - m_i)^\top, \quad (\text{der "within scatter"})$$

$$S_B = (m_1 - m_2)(m_1 - m_2)^\top \quad (\text{der "between scatter"}).$$

Die Lösung erhält man durch

$$w = S_W^{-1}(m_1 - m_2).$$

Das Programmskelett generiert einen einfachen Datensatz mit zwei Klassen und zeigt die herunterprojizierten Daten für die FDA-Richtung und eine andere (in **W1**) gespeichert an. In den beiden Plots darunter werden Histogramme für die herunterprojizierten Daten angezeigt, an denen man sehen kann, wie gut die Trennung der Daten ist.

Probiere auch verschiedene Richtungen für **W1** aus!

1. **S = scatter(X, M)** berechnet für eine  $d \times n$  Matrix  $X$  mit  $n$  Datenpunkten mit jeweils  $d$  Dimensionen und einem  $d$  dimensionalen Vektor  $m$  die Matrix

$$S = (X - m)(X - m)^\top.$$

**(7 Punkte)**

2. **W = fda(X, Y)** berechnet den optimalen Gewichtsvektor  $w$ . **(8 Punkte)**
3. **X = project(X, W)** berechnet die Daten, die auf den von  $W$  aufgespannten Raum projiziert wurden. Beachte:  $W$  muß nicht Länge 1 haben. **(7 Punkte)**
4. **show\_projected\_data(X, Y)** Plottet Histogramme für beide Klassen der eindimensionalen Daten  $X$ . (Hinweis: **hist** gibt die Histogrammdaten zurück, anstatt ein Histogramm anzuzeigen, wenn man Rückgabeargumente angibt) **(8 Punkte)**

---

**function sheet06**

```
% Generate some data
N = 100;
M1 = [1; 2];
M2 = [-5; 7];
X = [randn(2, N) + repmat(M1, 1, N), ...
      randn(2, N) + repmat(M2, 1, N)];
Y = [ones(1, N), -ones(1, N)];
```

```

% Show data
subplot(3,1,1);
plotdata(X, Y);

% Compute fda direction and another one and show
% the projected data
W = fda(X, Y)
W1 = [1; 0];

hold on;
plotdata(project(X, W), Y);
plotdata(project(X, W1), Y);
hold off;

% Histograms for the fda direction
subplot(3,1,2);
show_projected_data(W'*X, Y);

% Histograms for the other direction
subplot(3,1,3);
show_projected_data(W1'*X, Y);

% Plot two classes.
function plotdata(X, Y)
P = (Y == 1);
N = (Y == -1);

plot(X(1, P), X(2, P), 'r+', X(1, N), X(2, N), 'bo');

axis equal
grid on

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Your solutions below
%

% 1. Return scatter matrix  $S = (X - M)(X - M)'$ 
function S = scatter(X, M)
% ...

% 2. Compute FDA direction.
function W = fda(X, Y)
% ...

% 3. Project X to subspace spanned by W.
function X = project(X, W)
% ...

% 4. Plot histograms for both classes of a
% one-dimensional data set X.
function show_projected_data(X, Y)
% ...

```

---

Für Fragen zum Übungsblatte bitte in der Google Group <http://groups.google.com/group/ml-tu> registrieren und die Frage an die Mailingliste stellen.