

**Blatt 8**

Abgabe bis Montag, 7. Juni 2010, 13:00 Uhr bei Dr. Konrad Rieck (rieck@cs.tu-berlin.de)

In der Vorlesung wurde die RDE-Methode besprochen, um die effektive Dimensionalität im Kernmerkmalsraum zu bestimmen. Auf diesem Aufgabenblatt soll die Methode implementiert und einige Datensätze untersucht werden.

Ergänze die entsprechenden Funktionen im Programmskelett:

1. **(7 Punkte)** Schreibe eine Funktion `sincdata` mit Argumenten  $n$  und  $s$ , die  $n$  Punkte von folgenden Verteilungen generiert:

$$\begin{aligned} X_i &\sim \text{gleichverteilt von } -4, 4 \\ Y_i &\sim \sin(\pi X_i)/(\pi X_i) + s\varepsilon_i, \end{aligned}$$

wobei die  $\varepsilon_i$  normalverteilt sind.

2. **(7 Punkte)** Schreibe eine Funktion `sinedata` mit Argumenten  $n$  und  $k$ , die  $n$  Punkte von der Verteilung

$$\begin{aligned} X_i &\sim \text{gleichverteilt von } -\pi, \pi \\ Y_i &\sim \sin(kX_i) + 0,3\varepsilon_i. \end{aligned}$$

zieht, wobei die  $\varepsilon_i$  normalverteilt sind.

3. **(10 Punkte)** Implementiere `rde` basierend auf dem log-likelihood-Schätzer

$$\hat{d} = \operatorname{argmin}_{1 \leq d \leq \lfloor \frac{n}{2} \rfloor} \frac{d}{n} \log \frac{1}{d} \sum_{i=1}^d s_i^2 + \frac{n-d}{n} \log \frac{1}{n-d} \sum_{i=d+1}^n s_i^2,$$

wobei  $s = (u_1^T Y, \dots, u_n^T Y)$ ,  $u_1, \dots, u_n$  die Eigenvektoren der Kernmatrix sind (wie üblich sortiert nach absteigendem zugehörigem Eigenwert).

Zurückgegeben werden soll die effektive Dimensionalität und  $\hat{Y}$ , dass die herunterprojizierten Labels enthält:

$$\hat{Y} = \sum_{i=1}^{\hat{d}} u_i u_i^T Y.$$

4. **(6 Punkte)** Schreibe die Funktion `plot_fit`, die die Daten  $X, Y$  und die herunterprojizierten Labels  $\hat{Y}$  enthält.

---

```
function sheet08

%
% SINC DATA
%
figure(1)

% generate some data
[X, Y] = sincdata(100, 0.1);

% generate the kernel matrix
K = rbfkern(0.1, X);

% Run rde
[D, Yh] = rde(K, Y);
```

```

% plot the data
plot_fit(X, Y, Yh);
title(sprintf('sinc data set, effective dimensionality = %d', D));

%
% SINE DATA
%
figure(2)

% generate some data
[X, Y] = sinedata(100, 4, 0.1);

% generate the kernel matrix
K = rbfkern(0.1, X);

% Run rde
[D, Yh] = rde(K, Y);

% plot the data
plot_fit(X, Y, Yh);
title(sprintf('sine data set, effective dimensionality = %d', D));

function K = rbfkern(w, X)
N = size(X, 1);
XX = sum(X.*X, 2);
D = repmat(XX, 1, N) + repmat(XX', N, 1) - 2 * X * X';
K = exp(-D/(2*w));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Insert your solutions below
%

function [X, Y] = sincdata(N, noise)
% ...

function [X, Y] = sinedata(N, K, noise)
% ...

function [D, Yh] = rde(K, Y)
% ...

function plot_fit(X, Y, Yh)
% ...

```

---