

Blatt 5

Abgabe bis Montag, 17. Mai 2010, 13:00 Uhr bei Dr. Konrad Rieck.

Auf diesem Übungszettel sollen Kerne untersucht werden. Die zugrundeliegenden Datensätze sind einmal die “verrauschte Sinc-Funktion”

$$\begin{aligned}x &\sim \text{uniform in } [-10, 10] \\y &= \sin(x)/x + \varepsilon, \\ \varepsilon &\sim \mathcal{N}(0, 0,01),\end{aligned}$$

wobei $\mathcal{N}(m, v)$ die Gaussverteilung mit Mittelwert m und Varianz v ist.

Der zweite Datensatz ist ein Zweiklassenproblem bestehend aus zwei normalverteilten Klassen:

$$\begin{aligned}p(x|y = -1) &= \mathcal{N}((1, 2)^\top, I), \\p(x|y = 1) &= \mathcal{N}((-2, -3)^\top, I), \\p(y = 1) &= p(y = -1) = \frac{1}{2},\end{aligned}$$

wobei I die 2×2 Einheitsmatrix ist.

Zur Erinnerung: Die Kernel-PCA-Komponenten ausgewertet auf den Datenpunkten, auf denen auch die Zerlegung berechnet wurde, sind gerade die Eigenvektoren der Kernmatrix.

Ergänze die fehlenden Teil im Programmskelett

1. **(5 Punkte)** `[X, Y] = sincdata(N)` Erzeuge N Punkte des Sincfunktionsdatensatzes. Sortiere die X e, damit die Visualisierung später einfacher wird.
2. **(5 Punkte)** `[X, Y] = twoblobdata(N)` Erzeuge N Punkte des obigen Zweiklassenproblems.
3. **(5 Punkte)** `K = rbfkern(w, X, Y)` Berechnen die Kernelmatrix für alle Kombinationen von Zeile von X und Y . D.h.

$$K_{ij} = k(X_i, Y_i) = \exp\left(-\frac{\|X_i - Y_i\|^2}{2w}\right)$$

wobei X_i die i te Zeile von X ist. Verwende keine for-Schleifen! (Hinweis: Beachte, dass $\|x - y\|^2 = \|x\|^2 - 2x^\top y + \|y\|^2$ und schreibe diesen Ausdruck mit Matrizen so um, dass er direkt alle Abstände berechnet)

4. **(5 Punkte)** `plotkpca1d(K, X, D)` Plote die ersten D Kernel-PCA-Komponenten mit Hilfe der Kernmatrix K über den eindimensionalen Daten X .
5. **(5 Punkte)** `plotkpca2d(K, X, D)` Plote die ersten D Kernel-PCA-Komponenten mit Hilfe der Kernmatrix K über den zweidimensionalen Daten X . Verwende `subplot` um für jede Komponente einen eigenen Plot anzulegen. (siehe Beispielplo unten)
6. **(5 Punkte)** Variiere die Kernparameter und untersuche, wie sich die Komponenten für die beiden Datensätze verändern. Gib die entsprechenden Plots als png-Grafik für interessante Konstellationen ab.

```
function sheet05
```

```
%
% sinc function data set
%
[X, Y] = sincdata(100);
```

```

figure(1)
plot(X, Y, '.');
title('the data')

figure(2)
plotkpca1d(rbfkern(5, X, X), X, 5)
legend('1', '2', '3', '4', '5')
title('first 5 kPCA dimensions')
grid

%
% two blobs data set
%
[X, Y] = twoblobdata(500);
figure(3)
gscatter(X(:, 1), X(:, 2), Y);

figure(4)
w = 10;
plotkpca2d(rbfkern(w, X, X), X, 9)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% Your solutions below
%
function [X, Y] = sincdata(N)
% ...

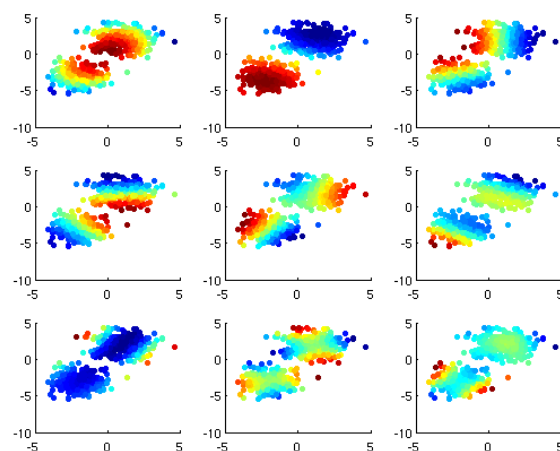
function [X, Y] = twoblobdata(N)
% ...

function K = rbfkern(w, X, Y)
% ...

function plotkpca1d(K, X, D)
% ...

function plotkpca2d(K, X, D)
% ...

```



Beispielplot für plotkpca2d.

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