

Informationsblatt

Ziel

Ziel dieses Praktikums ist es, die Implementierung von Maschinellen Lernalgorithmen und deren Anwendungen auf reale Daten zu üben. Behandelt werden die bekanntesten Algorithmen zur Dimensionsreduktion bzw. Visualisierung, Clustering, Regression, Klassifikation (inkl. Modellselektion) und Inferenz in Graphischen Modellen.

Voraussetzungen

- Grundkenntnisse in der Programmierung mit Matlab (z.B. durch Besuch des Kurses “Einführung in die computergestützte Datenanalyse mit Matlab”).
- Grundlagen des Maschinellen Lernens (z.B. durch die Vorlesung “Maschinelles Lernen I”) sind hilfreich, aber nicht unbedingt notwendig.
- Mathematik, insbesondere Wahrscheinlichkeitstheorie und Lineare Algebra auf Grundstudiumsniveau.

Ablauf

Die genauen Termine werden auf der Website bekannt gegeben. Der Ablauf ist wie folgt, pro Übungsblatt sind zwei oder drei Wochen Bearbeitungszeit vorgesehen.

Montag 10:00 Uhr (FR 3001)	Vorlesung, Ausgabe des Übungsblattes
Mittwoch 10:00 Uhr (FR 6043)	Vorstellung der Lösungen des letzten Übungsblattes, gemeinsame Sprechstunde zum aktuellen Übungsblatt
Mittwoch 10:00 Uhr (FR 6043)	Gemeinsame Sprechstunde zum aktuellen Übungsblatt
...	...
Montag 10:00	Abgabeschluss

Jeder Teilnehmer muss einmal seine Lösung am Mittwoch vorstellen. Eine Liste zur Wahl des Themas (inkl. vorläufigem Termin) wird in Kürze veröffentlicht.

Vorläufige Übersicht der Themen

- Übungsblatt 1: Matlab
- Übungsblatt 2: Dimensionsreduktion bzw. unsupervised learning (PCA, Isomap, LLE)
- Übungsblatt 3: Clustering (k -means, EM, Hierarchical Clustering)
- Übungsblatt 4: Einführung in die Klassifikation und Modellselektion (Kernel-ridge-regression, Decision Trees, Boosting, Cross Validation)
- Übungsblatt 5: Support-Vector-Maschinen
- Übungsblatt 6: Inferenz in Graphischen Modellen (factor graphs, loopy belief propagation)

Technik

Programmierrichtlinien

Bitte beachtet für Eure Abgaben die folgenden Richtlinien:

- Name und Signatur der Aufgaben werden in der Regel fest vorgegeben. Bitte haltet Euch daran.
- Jede Datei muß Euren Namen enthalten.
- Der Code muß eigenständig in der IRB-Umgebung lauffähig sein (cs-Netz).
- Jede Funktion muß kommentiert werden. Bitte haltet Euch dabei an folgendes Schema:

```
function [D1, D2, td] = Aufgabe1(X)
% Aufgabe1 - compute distance matrix
%
% usage
%     [D1, D2, td] = Aufgabe1(X)
%
% input
%     X : (d,n)-matrix of column vectors
%
% output
%     D1 : (n,n)-matrix of L_2 distances, computed by for loop
%     D2 : (n,n)-matrix of L_2 distances, computed by matrix algebra
%     td : run-time difference between computation of D1 and D2
%
% description
%     Aufgabe1 computes the pair-wise L_2 distances between the column
%     vectors of X using two different algorithms: D1 by explicit
%     for-loops, D2 by matrix algebra. td reports the difference in
%     running time between the two implementations.
%
% author
%     Paul Buenau, buenau@cs.tu-berlin.de
```

Matlab

Matlab kann auf den Rechnern des IRB (pepino, bolero, etc.) mit dem Kommando

~ml/bin/matlab

gestartet werden.

Leistungsnachweis

Die Endnote setzt sich zusammen aus den Leistungen in den Uebungsblaettern (50%) und einer muendlichen Abschlusspruefung (50%). Die Übungsblätter dürfen nicht in Gruppen bearbeitet werden.

Kontakt

Dr. Mikio Braun, Raum FR 6068, Telefon 314-78627, mikio@cs.tu-berlin.de, Sprechstunde jederzeit nach Vereinbarung (ab 10.5.)

Paul von Buenau, Raum FR 6059, Telefon 314-78628, buenau@cs.tu-berlin.de, Sprechstunde jederzeit nach Vereinbarung

Sekretariat FR 6-9: Andrea Gerdes, Raum FR 6052, Telefon 314-78621, gerdes@cs.tu-berlin.de

Webseite: <http://ml01.zrz.tu-berlin.de:8000/wiki/Main/SS09.MLPraktikum>

Aufgabenblaetter, Loesungen, Daten, Punktzahlen, Termine und ggf. Literaturhinweise werden auf der Webseite veroeffentlicht.