Maschinelles Lernen

8. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade
Paul Prasse

WS09/10

Ausgabe am: 05.01.09
Besprechung am: 12.01.10

Aufgabe 1 (1/4 Punkt):
Betrachten wir die Trainingsbeispiele aus der letzten Übung: y_i ist das Klassenlabel, x_{1i} und x_{2i} sind die Attribute.

i	1	2	3	4	5	6
y_i	+1	+1	+1	–1	–1	–1
x_{1i}	1,0	1,3	2,5	4,2	5,1	3,7
x_{2i}	3,4	4,2	3,7	1,7	2,5	3,1

Der Perzeptron-Algorithmus (ohne expliziten Bias b) ist wie folgt gegeben:

Algorithm 1 Perzeptron

Require: n Beispiele $(\mathbf{x}_i, y_i)$
$\mathbf{w} := \mathbf{0}$
repeat
  $updates := 0$
  for $i = 1$ to n **do**
    if $y_i \mathbf{w}^\mathsf{T} \mathbf{x}_i \leq 0$ **then**
      $\mathbf{w} := \mathbf{w} + y_i \mathbf{x}_i$
      $updates := updates + 1$
    end if
  end for
until $updates = 0$

(a) Zeigen sie, dass ein Update $\mathbf{w}' := \mathbf{w} + y_i \mathbf{x}_i$ tatsächlich den Zielfunktionswert $\mathbf{w}'^\mathsf{T} \mathbf{x}_i$ in die gewünschte Richtung verändert, d.h. $y_i \mathbf{w}'^\mathsf{T} \mathbf{x}_i > y_i \mathbf{w}^\mathsf{T} \mathbf{x}_i$.

(b) Simulieren sie das Training eines Perzeptrons $f(x) = \mathbf{w}^\mathsf{T} \mathbf{x}$ von Hand. Laufen sie in jeder Iteration des Algorithmus von links nach rechts über die Tabelle. Klassifizieren sie Testbeispiel $[3,5; 3,5]^\mathsf{T}$ unter Verwendung der Klassifikationsfunktion $sign(f(x))$. Tipp: Um Zeit zu sparen, brauchen sie den Zielfunktionswert nicht für jedes Beispiel explizit ausrechen. Oft sieht man schon an den Vorzeichen und Größenordungen der Gewichtsvektoreinträge, dass ein Beispiel richtig klassifiziert wird und kein Update nötig ist.

(c) Nehmen sie an, sie hätten ein zusätzliches positives Trainingsbeispiel $[5,0; 1,0]^\mathsf{T}$. Sie möchten nun wieder ein Perzeptron trainieren. Stellen sie die Trainingsdaten grafisch in einem zweidimensionalen Diagramm dar und überlegen sie sich ohne ausprobieren, welcher Klasse das obige Testbeispiel nun zugeordnet werden würde.

Aufgabe 2 (1/4 Punkt):
Das Margin-Perzeptron führt ein Update des Gewichtsvektors durch wenn

$$y_i \frac{\mathbf{w}^\mathsf{T} \mathbf{x}_i}{\|\mathbf{w}\|} \leq \delta.$$

(a) Welche Verlustfunktion minimiert das Margin-Perzeptron?

(b) Zu Beginn des Trainings ist der maximale Margin δ_{max} i.A. unbekannt. Angenommen wir wählen $\delta > \delta_{max}$. Welche Auswirkung hat diese Wahl auf das Ergebnis des Margin-Perzeptrons?

Aufgabe 3 (1/4 Punkt):
In der Vorlesung wurde das Optimierungskriterium der *Support Vector Machine* vorgestellt.

(a) Wir betrachten zunächst die *Hard Margin SVM*. Begründen sie warum das Minimieren der Länge des Gewichtsvektors unter der Nebenbedingung $y_i \mathbf{w}^\mathsf{T} \mathbf{x}_i \geq 1, \forall i$, äquivalent zur Maximierung des geometrischen Margins ist.

(b) Was sind die Support-Vektoren?

(c) Worin besteht der Unterschied zw. *Hard* und *Soft Margin SVM*? Geben sie die Verlustfunktionen und die Regularisierer beider Verfahren an.

Aufgabe 4 (1/4 Punkt):
Bisher haben wir im Zusammenhang mit dem Perzeptron, der SVM usw. die Daten immer als Punkte im Raum aufgefasst welche wir durch eine Ebene – im zweidimensionalen eine Gerade – trennen möchten (im unteren Bild rechts). Eine äquivalente Darstellung bildet der *konjugierte Vektorraum*. In diesem wird jeder Datenpunkt als Normalenvektor $y_i \frac{\mathbf{x}_i}{\|\mathbf{x}_i\|}$ einer Ebene aufgefasst. Der Normalenvektor $\mathbf{w}$ einer Trennebenen (mit Länge 1) wird hingegen als Punkt dargestellt (im unteren Bild links). Wo befindet sich in dieser Darstellung $\mathbf{w}_{SVM}$ (Lösung der Hard Margin SVM) und wo liegt der Bayes Point $\mathbf{w}_{BP}$ (Lösung der Bayes Point Machine)?

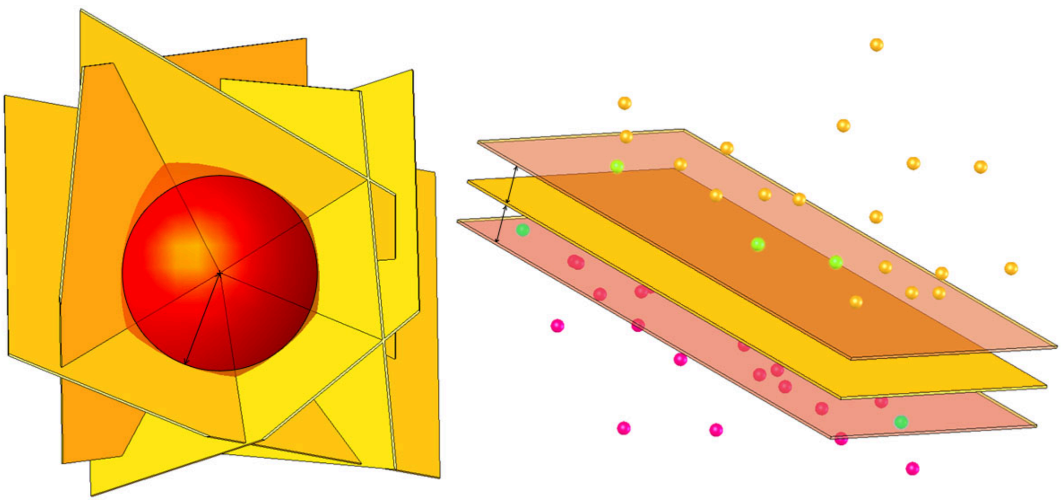

Abbildung 1: SVM im Vektorraum (rechts) und im konjugierten Vektorraum (links)