Maschinelles Lernen

9. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade
Silvia Makowski

WS11/12

Ausgabe am: 04.01.12
Besprechung am: 10.01.12

Aufgabe 1 *Support Vector Machine*

In der Vorlesung wurden verschiedene Optimierungskriterium für lineare Klassifikatoren vorgestellt.

(a) Das Margin-Perzeptron führt ein Update des Gewichtsvektors durch wenn

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

Welche Verlustfunktion minimiert das Margin-Perzeptron?

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

(c) Diskutieren Sie woher der Name *support vector* stammen könnte? Welchen Einfluss haben Beispiele mit $y_i \mathbf{x}_i^\mathsf{T}\mathbf{w}^* > 1$ auf die optimale Lösung $\mathbf{w}^*$?

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

Aufgabe 2 *Bayes Point*

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)?

Aufgabe 3 *Logistische Regression*

Die *Logistische Regression* – anders als der Name vermuten lässt ein Klassifikationsverfahren – basiert auf der Idee, die Klassenwahrscheinlichkeiten

$$p(y_i = +1|\mathbf{x}_i) = 1 - p(y_i = -1|\mathbf{x}_i)$$

mit Hilfe einer Sigmoidfunktion $\sigma(f(\mathbf{x}_i))$ zu modellieren. Oft wird dabei die *logistische Funktion* $\sigma(a) = \frac{1}{1+e^{-a}}$ verwendet.

(a) Angenommen wir modellieren die Wahrscheinlichkeit $p(y_i = +1|\mathbf{x}_i; \mathbf{w}) = \sigma(\mathbf{w}^\mathsf{T}\mathbf{x}_i)$. Zeigen Sie, dass für die Gegenwahrscheinlichkeit $p(y_i = -1|\mathbf{x}_i; \mathbf{w})$ gilt:

$$p(y_i = -1|\mathbf{x}_i, \mathbf{w}) = \sigma(-\mathbf{w}^\mathsf{T}\mathbf{x}_i).$$

(b) Aus Aufgabe (a) wissen wir, dass wir die Klassenwahrscheinlichkeiten mit nur einer Funktion modellieren können: $p(y_i|\mathbf{x}_i;\mathbf{w}) = \sigma(y_i\mathbf{w}^\mathsf{T}\mathbf{x}_i)$. Leiten Sie einen ML-Schätzer für $\mathbf{w}$ her (nur das Optimierungskriterium). Nehmen Sie dabei an, dass die Datenpunkte unabhängig voneinander gezogen wurden, d.h.

$$p(y_1,\ldots,y_n|\mathbf{x}_1,\ldots,\mathbf{x}_n;\mathbf{w}) = \prod_{i=1}^{n} p(y_i|\mathbf{x}_i;\mathbf{w}).$$

(c) Wir nehmen an, dass der Prior des Gewichtsvektors $p(\mathbf{w})$ normalverteilt ist mit $p(\mathbf{w}) = N(\mathbf{w}^*, \tau^2\mathbf{I})$, wobei $\mathbf{w}^*$ z.B. ein altes Modell ist. Leiten sie den MAP-Schätzer (nur das Optimierungskriterium) der Logistischen Regression her und geben Sie die Verlustfunktion und den Regularisierer an.

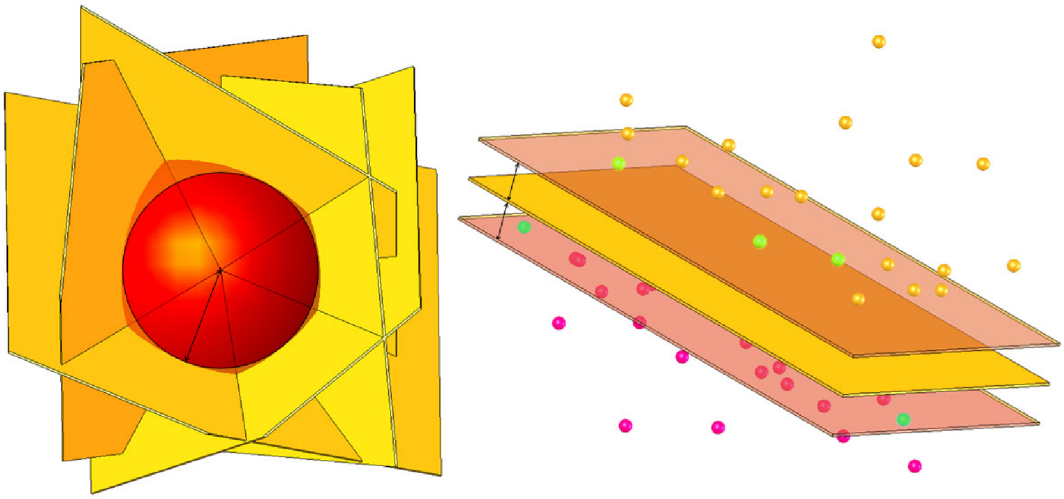

Abbildung 1: SVM im Vektorraum (rechts) und im konjugierten Vektorraum (links)