

Maschinelles Lernen

10. Übung

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

WS09/10

Ausgabe am: 19.01.10
Besprechung am: 26.01.10

Aufgabe 1 (1/3 Punkt):

Neben expliziten Abbildungen $\phi : \mathbf{x} \rightarrow \mathbf{x}'$, welche die ursprünglichen Daten in einen neuen Raum mappen, existieren auch Kernel-Funktionen.

- (a) Was ist der Unterschied zw. einer Basis-Funktion $\phi(\mathbf{x})$ und einer Kernel-Funktion $k(\mathbf{x}, \mathbf{x}')$? Welche Vor-/Nachteile hat eine Kernel-Funktion gegenüber der Verwendung eines Mappings ϕ ?
- (b) Wie könnte ein Graph-Kernel und wie ein Text-Kernel aussehen?
- (c) Was ist eine Kernel-Matrix und welche Eigenschaften hat sie?
- (d) Was besagt das Representer-Theorem?

Aufgabe 2 (1/3 Punkt):

Die Multiklassen-SVM kann durch folgendes Optimierungsproblem dargestellt werden:

$$\min_{\mathbf{w}} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

unter den Nebenbedingungen:

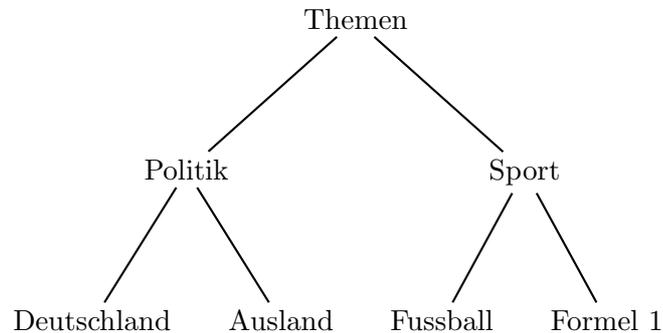
$$\begin{aligned} \forall i \in \{1, \dots, n\} \forall y \neq y_i : \langle \mathbf{w}, \Phi(\mathbf{x}_i, y_i) \rangle &\geq \langle \mathbf{w}, \Phi(\mathbf{x}_i, y) \rangle + 1 - \xi_i \\ \forall i \in \{1, \dots, n\} : \xi_i &\geq 0. \end{aligned}$$

Zeigen Sie, dass die (binäre) SVM ein Spezialfall dieser Formulierung ist. Vergleichen Sie die optimalen Gewichtsvektoren beider Lösungen.

Aufgabe 3 (1/3 Punkt):

Geben Sie für die Probleme mit strukturierten Ausgabebereichen die gemeinsame Repräsentation von Ein- und Ausgabe $\Phi(x, y)$ an.

- a) Sie wollen einen Klassifikator lernen, der Texte anhand ihrer Überschriften einem Thema zuordnet. Diese sind durch die folgende Hierarchie gegeben:



Geben Sie $\Phi(x, y)$ für die folgenden Beispiele an:

- i) “Zico getroffen, Tevez trifft doppelt” → Thema: Fussball
- ii) “Merkel trifft Stoiber” → Thema: Deutschland

Welche Abschnitte des gelernten Gewichtsvektor kann man unterscheiden? Diskutieren Sie Möglichkeiten, einen Lernalgorithmus so anzupassen, dass er einen Klassifikator bevorzugt, der möglichst lange richtige Pfade vorhersagt.

- b) Sie wollen einen Klassifikator lernen, der für Internetbenutzer das Alter und den Familienstand aus den Themen der von ihnen besuchten Webseiten vorhersagt. Dabei werden die diskreten Altersabschnitte $Y_1 = \{< 12; 12 - 25; 26 - 45; > 45\}$ und Familienstände $Y_2 = \{ledig; verheiratet; verwitwet\}$ unterschieden. Geben Sie $\Phi(x, y)$ für die folgenden Beispiele an:

- i) Thema: Fussball → Alter: 12-25, Familienstand: ledig
- ii) Thema: Deutschland → Alter: > 45, Familienstand: verwitwet

Welche Abschnitte des gelernten Gewichtsvektor kann man unterscheiden? Was ist der Vorteil dieser kollektiven Klassifikation gegenüber einer unabhängigen Vorhersage von Y_1 und Y_2 mit Hilfe einer Multiklassen-SVM?