Einführung in das maschinelle Lernen

Christian Pölitz

23. April 2014

technische universität
dortmund

Support Vector Maschinen

Eine kleine Wiederholung vom letzten Mal.

Maximal Margin Classifier

Für einen gegebenen Datensatz $S = \{(x_1, y_1), \cdots, (x_n, y_n)\}$, suchen wir eine lineare Funktion $g(x) =< \beta, \phi(x) > +\beta_0$, die für ein gegebenes x, ein korrespondierendes y vorhersagt.

Klassifikation

$y_i \in \{-1, 1\}$ und wir wollen $g(x_i) \leq -\gamma$ wenn $y_i = -1$ und $g(x_i) > \gamma$ sonst. Generell: $y_i \cdot (< \beta, \phi(x_i) > +\beta_0) \geq \gamma$

Margin

Der Margin für einen Datensatz S und eine lineare Funktion $g(x)$ ist definiert durch: $m(S,g) = \min y_i \cdot g(x_i) \geq \gamma$.

Bild SVM

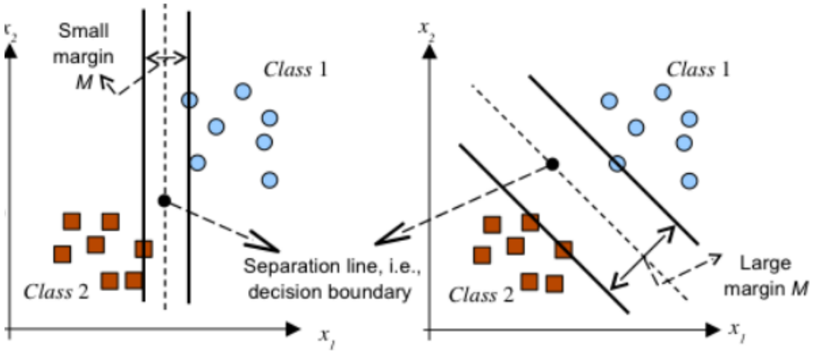

Abbildung: $http : //yottamine.com/machine - learning - svm$

Formalisierung als Optimierungsaufgabe

$$
\begin{aligned}
max \quad & \gamma & (1)\\
s.t. \quad & y_i \cdot (< \beta, \phi(x_i) > + \beta_0) \geq \gamma & (2)\\
& ||\beta||^2 = 1 & (3)
\end{aligned}
$$

technische universität dortmund

Formalisierung als Optimierungsaufgabe als Minimierungsproblem

$$\begin{aligned}
min \quad & -\gamma && (4) \\
s.t. \quad & -y_i \cdot (< \beta, \phi(x_i) > + \beta_0) + \gamma \leq 0 && (5) \\
& ||\beta||^2 = 1 && (6)
\end{aligned}$$

Lagrangian (Primales Problem)

$$L_P(\beta, \beta_0, \alpha, \lambda) =$$
$$-\gamma - \sum_i \alpha_i \cdot (y_i \cdot (< \beta, \phi(x_i) > +b) + \gamma) + \lambda \cdot (||\beta||^2 - 1)$$

Partielle Ableitungen

$$\frac{\partial L_P(\beta,\beta_0,\alpha,\lambda)}{\partial\beta} = \beta - \frac{1}{2}\cdot\lambda\cdot\sum_i \alpha_i\cdot y_i\cdot\phi(x_i) = 0$$
$$\frac{\partial L_P(\beta,\beta_0,\alpha,\lambda)}{\partial\beta_0} = -1 + \sum_i \alpha_i = 0$$
$$\frac{\partial L_P(\beta,\beta_0,\alpha,\lambda)}{\partial\gamma} = -\sum_i \alpha_i\cdot y_i = 0$$

technische universität dortmund

Lagrangian (Duales Problem)

- $L_P(\alpha) = -\sum_{i,j} \alpha_i \cdot \alpha_j \cdot y_i \cdot y_j \cdot \phi(x)^T \phi(x) \rightarrow max$
- $\sum_i \alpha_i \cdot y_i = 0$
- $\alpha_i \geq 0$

Form der Lösung

$\beta = \frac{1}{2} \cdot \lambda \cdot \sum_i \alpha_i \cdot y_i \cdot x_i$ und damit ist die Lösung eine lineares Funktional der Form $\sum_i a_i \cdot \phi(x_i)$.

Kernel Methoden

- Einbettung der Daten in einem Vectorraum V
- In diesem Vektorraum werden lineare Beziehungen und trennende Funktionale gesucht
- Algorithmen arbeiten nicht auf den Elementen von V, sondern mit paarweisen inneren Produkten $< .,. >$

Bild Kernel Methoden

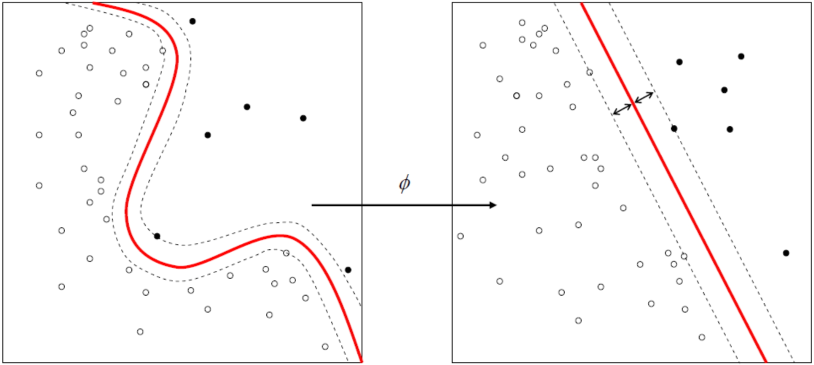

Abbildung:
http : //commons.wikimedia.org/wiki/File : Kernel_Machine.png

Kernel

Inneres Produkt in V, das von einer Abbildung ϕ kommt.

- $k(x, y) =< \phi(x), \phi(y) >$
- $\phi : x \to \phi(x) \in V$

Beispiel

$$\phi((x_1, x_2)) = (x_1^2, x_2^2, \sqrt{2} \cdot x_1 \cdot x_2) \in V = \Re^3$$

Beispiel

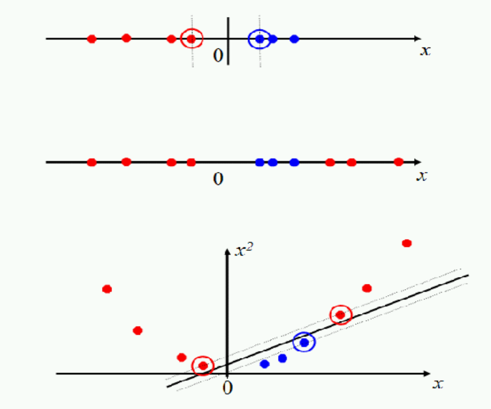

Abbildung: http://nlp.stanford.edu/IR-book/html/htmledition/nonlinear-svms-1.html

Beispiel

Zwei Teilmengen A_1 und A_2 einer Menge Ω $k(A_1, A_2) = 2^{|A_1 \cap A_2|}$ ist ein inneres Produkt der Abbildung:

$$\phi(A)_U = \begin{cases} 1, U \subseteq A \\ 0, \text{otherwise} \end{cases} \tag{7}$$

U sind Teilmengen von Ω, also Elemente der Potentmenge von Ω.

Hilbertraum

Ein vollständiger normierter Vektorraum mit einem inneren Produkt $< ., . >$ ist ein Hilbertraum.

- Vollständig heißt, dass alle Cauchy Folgen konvergieren
- Ein Vektorraum ist eine Menge, die bezüglich Addition und Skalarmultiplikation (mit Skalaren eines bestimmten Körpers) abgeschlossen ist

Inneres Produkt

- Symmetrische Bilinearform: $< \alpha \cdot x, \beta \cdot y >= \alpha \cdot \overline{\beta} \cdot < x, y >$
- Positiv definit: $< x, x > \geq 0$ und $< x, x >= 0 \Longleftrightarrow x = 0$
- Norminduzierend: $||x|| = \sqrt{< x, x >}$
- Standard Inner Prduct: $< x, y >= \sum x_i \cdot y_i$

Winkel

- $\cos\theta = \frac{<x,y>}{||x||\cdot||y||}$
- $\theta = 0 \Rightarrow \cos\theta = 1$ und $< x, y >= ||x|| \cdot ||y||$
- $\theta = \frac{\pi}{2} \Rightarrow \cos\theta = 0$ und $< x, y >= 0$ Orthogonalität

Darstellung von Riesz

Es existiert eine isometrischer Isomorohismus von einem Hilbertraum V in seinen (stetigen) Dualraum V'. D.h. jedes Element $f_v \in V$ genügt $f_v(x) =< x, v >$. Die Punktauswertung ist ein stetiges lineares Funktional. Unsere Abbildungen sind also innere Produkte, die in einem Argument fest sind. $\phi(x) = k(x, .)$.

Beispiel

Die Menge aller Folgen reellwertiger Zahlen $l_2(\mathbb{N}) = \{(x_1, \cdots, x_n, \cdots) | \sum x_i^2 < \infty\}$ mit dem inneren Produkt $< x, y > = \sum x_i \cdot y_i$ ist eine Hilberraum.

Beispiel

Die Menge der quadratintegrierbaren Funktionen $L_2(X) = \{f | \int_X f(x)^2 \cdot dx < \infty\}$ mit dem inneren Produkt $< x, y >= \int_X f(x) \cdot g(x) \cdot dx$ ist ein Hilbertraum.

Gram Matrix

Alle paarweise inneren Produkte für eine Menge S können in einer Matrix $G_{i,j} =< \phi(x_i), \phi(x_j) >= k(x, y)$ zusammengefasst werden. Diese Matrix ist symmetrisch und positiv definit.

Feature Raum F

Für die SVM suchen wir ein lineares Funktional $\beta = \sum \beta_i \cdot \phi(x_i)$. Der Raum all dieser Funktionale ist ein Hilberraum (dessen Vervollständigung).

$$F = \{\sum b_i \cdot k(x_i, .)\}$$

Inneres Produkt in F

Sei $f = \sum \alpha_i \cdot k(x_i, .) \in F$ und $g = \sum \beta_i \cdot k(x_i, .) \in F$. Wir definiere follgendes innere Produkt:

$$< f, g >= \sum \sum a_i \cdot b_i \cdot k(x_i, x_j) = \sum a_i \cdot g(x_i) = \sum b_i \cdot f(x_i)$$

Reproducing Property

$$< f, k(x,.) >= \sum a_i \cdot k(x_i,.) = f(x)$$

Die Abbildungen

$\phi : X \rightarrow \Re$, X muss kein Menge reeller Zahlen sein. Wir können auch komplexe Strukturen, wie Parse Bäume oder Sequenzen von Wörtern sein.

Abschlusseigenschaften

- Addition
- Skalarmultiplikation
- Multiplikation
- Verknüpfung