Blatt 8

Montag, 25. Juni 2018 15:43

Aufgabe 4

a) zu zeigen: $k: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $k(x,y) := \sum_{i=1}^{n} x_i y_i$ ist positiv semi definiter Kernel

Beweis

1) Symmetry

$$k(x,y) = \sum_{n=1}^{m} x_n \cdot y_n = \sum_{n=1}^{m} y_n \cdot x_n = k(y,x) \quad \blacksquare$$

2) zu zeigen: $\forall c_1, \dots, c_m \in \mathbb{R}: \sum_{i,j=1}^{m} c_i c_j k(x_i, x_j) \geq 0$

mit $k(x,y) = \sum_{k=1}^{m} x_k \cdot y_k$

Beweis

$$\sum_{i,j=1}^{m} c_i c_j \left(\sum_{k=1}^{m} x_{i_k} \cdot x_{j_k}\right)$$

$$= \sum_{i,j=1}^{m} c_i c_j \, k(x_i, x_j)$$

$$= (c_1, \dots, c_m) \cdot \begin{pmatrix} k(x_1,x_1) & \cdots & k(x_1,x_m) \\ \vdots & \ddots & \vdots \\ k(x_m,x_1) & \cdots & k(x_m,x_m) \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ \vdots \\ c_m \end{pmatrix} = c^T K c$$

Angenommen: $K = B^T \cdot B$ dann $c^T K c = c^T B^T B c = (B \cdot c)^T \cdot Bc = \|Bc\|^2 \geq 0$

Beweis für $K = B^T B$:

$$\begin{pmatrix} k(x_1,x_1) & \cdots & k(x_1,x_m) \\ \vdots & \ddots & \vdots \\ k(x_m,x_1) & \cdots & k(x_m,x_m) \end{pmatrix} = \begin{pmatrix} x_1^T x_1 & \cdots & x_1^T \cdot x_m \\ \vdots & \ddots & \vdots \\ x_m^T x_1 & \cdots & x_m^T x_m \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_m^T \end{pmatrix} \cdot (x_1 \dots x_m) \quad \blacksquare$$

Mit 1) und 2) ist gezeigt, dass $k(x,y) := \sum_{i=1}^{n} x_i y_i$ ein positiv semi-definierter Kernel ist. $\blacksquare$

b) zu zeigen:

b) <u>zu zeigen:</u> $k: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $k(x,y) := \left(\sum_{i=1}^n x_i y_i\right)^2$ ist positiv semi definiter Kernel

<u>Beweis</u>

1) Symmetry

$$k(x,y) = \left(\sum_{i=1}^n x_i y_i\right)^2 \overset{a)}{=} \left(\sum_{i=1}^n y_i x_i\right)^2 = k(y,x) \quad \blacksquare$$

<u>zu zeigen:</u>

2) $\forall c_1, \dots, c_m \in \mathbb{R}: \sum_{i,j=1}^m c_i c_j \, k(x_i, x_j) \geq 0$

mit $\sum_{i,j=1}^m c_i c_j \left(\sum_{k=1}^n x_{ik} x_{jk}\right)^2 \geq 0$

c) <u>zu zeigen:</u> $k(x,y)^d$ ist ein Kernel.

<u>Beweis:</u> $k(x,y)^d = \prod_{i=1}^d k(x,y)$

$\Rightarrow$ durch den 2. Satz von dem Übungsblatt (Produkt von Kernel = Kernel) ist gezeigt, dass $k(x,y)^d$ ein Kernel ist $\blacksquare$

d) <u>zu zeigen:</u> $\sum_{i=0}^{i} c_i \cdot k^i$ ist ein Kernel.

<u>Beweis:</u> k_1 ist ein Kernel $\Rightarrow \langle \phi(x), \phi(y) = k_1(x,y)$

1) $$\begin{aligned} a \cdot k_1(x,y) &= a \cdot \langle \phi(x), \phi(y) \rangle \\ &= \sqrt{a} \cdot \sqrt{a} \cdot \langle \phi(x), \phi(y) \rangle \\ &= \langle \sqrt{a} \cdot \phi(x), \sqrt{a} \cdot \phi(y) \rangle \\ &= \langle f(x), f(y) \rangle \quad \text{mit } f(x) := \sqrt{a} \cdot \phi(x) \\ &= k_2(x,y) \end{aligned}$$

In Aufgabe 4c haben wir Bewiesen, dass $k(x,y)^i$ ein Kernel ist.
In 1) haben wir gezeigt $a \cdot k(x,y)^i$ ein Kernel ist.
Die Summe $\sum_{i=0}^n$ über Kernel ergibt aus dem 1. Satz von dem Übungsblatt einen Kernel.
$\Rightarrow$ Somit ist bewiesen dass $\sum^n c_i \cdot k(x,y)^i$ ein Kernel ist

Übungsblatt $\overset{i=0}{\text{ohen}}$ Uernel. $\cup$

$\Rightarrow$ Somit ist bewiesen, dass $\sum_{i=0}^{n} c_i \cdot k(x,y)^i$ ein Uernel ist. ■

e) zu zeigen: $(x,y) \mapsto f(x)\, k(x,y)\, f(y)$ ist ein Uernel

Beweis:
$$\begin{aligned} f(x)\, k(x,y)\, f(y) &= f(x) \langle \phi(x), \phi(y) \rangle f(y) \\ &= \langle f(x)\phi(x), f(y)\phi(y) \rangle \quad \text{mit } a(x) =: f(x) \cdot \phi(x) \\ &= \langle a(x), a(y) \rangle \\ &= k(x,y) \end{aligned}$$ ■

Aufgabe 1

zu zeigen: 2 Datenpunkte genügen um die Hyperebene mit maximaler Margin zu finden

Beweis: Gegeben bei einer SVM ist das folgende Minimierungsproblem. Da nur 2 Punkte (mit unterschiedlicher Klasse) im Set enthalten sind, soll die Margin für diese Punkte genau 1 bzw. -1 sein

$$\arg\min \frac{1}{2} \|\omega\|^2$$

$$\omega^T x_1 + b = 1 \qquad (1)$$
$$\omega^T x_2 + b = -1 \qquad (2)$$

Die dazugehörige Lagrangefunktion lautet

$$L(\omega, b, \lambda_1, \lambda_2) = \frac{1}{2} \|\omega\|^2 + \lambda_1 (\omega^T x_1 + b - 1) + \lambda_2 (\omega^T x_2 + b + 1)$$

Bilde nun die dazugehörige Ableitungen:

$$\frac{\partial L(\omega, b, \lambda_1, \lambda_2)}{\partial \omega} = \omega + \lambda_1 x_1 + \lambda_2 x_2 = 0 \qquad (3)$$

$$\frac{\partial L(\omega, b, \lambda_1, \lambda_2)}{\partial b} = \lambda_1 + \lambda_2 = 0 \Rightarrow \lambda_1 = \lambda_2 \qquad (4)$$

Durch Einsetzen von (3) in (4) erhält man:

$$\omega + \lambda_1 (x_1 - x_2) = 0 \Leftrightarrow \omega = -\lambda_1 (x_1 - x_2) \qquad (5)$$

Durch Addition von (1) und (2) erhält man:

$$\omega^T (x_1 + x_2) + 2b = 0 \Leftrightarrow 2b = -\omega^T (x_1 + x_2)$$

$\Rightarrow$ Man sieht, dass λ_1 in (3) nicht genauer bestimmt werden kann. λ_1 ist jedoch nur ein Skalar. λ_1 ändert somit nicht die Ordnung von ω. D.h. die Lage der Hyperline ist komplett bestimmt, nur die Skalierung nicht. ■

Aufgabe 2

zu zeigen: 2 Datenpunkte genügen um die Hyperebene mit maximaler Margin zu finden und $\gamma > 0$ verändert das Ergebnis nicht, sodass $y_n(\omega^T \phi(x_n) + b) \geq \gamma$ mit $n = 1, \dots, N$ $\gamma > 0$

Beweis: Gegeben bei einer SVM ist das folgende Minimierungsproblem. Da nur 2 Punkte (mit unterschiedlicher Klasse) im Set enthalten sind, soll die Margin für diese Punkte genau γ bzw. $-\gamma$ sein

$$\arg\min \frac{1}{2} \|\omega\|^2$$

$$\omega^T x_1 + b = \gamma \quad (1)$$
$$\omega^T x_2 + b = -\gamma \quad (2) \qquad \text{mit } \gamma > 0$$

Die dazugehörige Lagrangefunktion lautet

$$L(\omega, b, \lambda_1, \lambda_2) = \frac{1}{2} \|\omega\|^2 + \lambda_1 (\omega^T x_1 + b - \gamma) - \lambda_2 (\omega^T x_2 + b + \gamma)$$

Bilde nun die dazugehörige Ableitungen:

$$\frac{\partial L(\omega, b, \lambda_1, \lambda_2)}{\partial \omega} = \omega + \lambda_1 x_1 + \lambda_2 x_2 = 0 \quad (3)$$

$$\frac{\partial L(\omega, b, \lambda_1, \lambda_2)}{\partial b} = \lambda_1 + \lambda_2 = 0 \Rightarrow \lambda_1 = \lambda_2 \quad (4)$$

Durch Einsetzen von (3) in (4) erhält man:

$$\omega + \lambda_1 (x_1 - x_2) = 0 \Leftrightarrow \omega = -\lambda_1 (x_1 - x_2) \quad (5)$$

Durch Addition von (1) und (2) erhält man:

$$\omega^T (x_1 + x_2) + 2b = \gamma + (-\gamma) = 0 \Leftrightarrow 2b = -\omega^T (x_1 + x_2)$$

$\Rightarrow$ Man sieht, dass γ keinen Einfluss auf das Ergebnis hat. $\blacksquare$