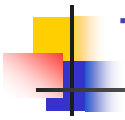
Kerndesign für Vektorraumdaten

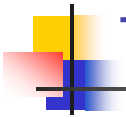
Andreas Werber

24. Januar 2003



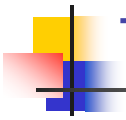
Themen

- Kernel-Trick



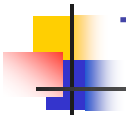
Themen

- Kernel-Trick
- Anforderungen an Kerne



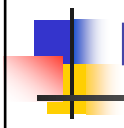
Themen

- Kernel-Trick
- Anforderungen an Kerne
- Konstruktionsmöglichkeiten



Themen

- Kernel-Trick
- Anforderungen an Kerne
- Konstruktionsmöglichkeiten
- Problemspezifische Anwendungen



Kernel-Trick (Wiederholung)

Kernel-Trick

- Trainingsdaten treten niemals einzeln auf, sondern stets in Form von Skalarprodukten zwischen Paaren von Beispielen

Kernel-Trick

- Trainingsdaten treten niemals einzeln auf, sondern stets in Form von Skalarprodukten zwischen Paaren von Beispielen
- Gram-Matrix = Kern-Matrix*

$$\mathbf{G} = \mathbf{K} = \left(< \mathbf{x_i} \cdot \mathbf{x_j} > \right)_{i,j=1}^{n,n}$$

  - Die Gram-Matrix entsteht durch Skalarprodukte zwischen einer Menge von Vektoren, ist also über einer Menge von Punkten definiert
  - Die Kern-Matrix entsteht durch Auswertungen einer Kernfunktion auf einer Menge von Objekten

Kernel-Trick

- Die Anzahl der Parameter einer linearen Maschine im Merkmalsraum (bzw. einer nicht-linearen Maschine im Ursprungsraum) ist abhängig von der Dimension des Eingaberaumes

$$f(\mathbf{x}) = \sum_{i=1}^{N} \omega_i \cdot \varphi_i(\mathbf{x}) + b$$

$\boldsymbol{\varphi} : X \rightarrow F$ ist eine nicht-lineare Abbildung in den Merkmalsraum

Kernel-Trick

- Die Anzahl der Parameter einer linearen Maschine in der dualen Darstellung ist unabhängig von der Anzahl der Attribute (Parameterzahl abhängig von der Anzahl der verwendeten Samples)

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot < \boldsymbol{\varphi}(\mathbf{x_i}) \cdot \boldsymbol{\varphi}(\mathbf{x}) > + b$$

Kernel-Trick

- Ersetzen des Skalarproduktes $<\varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x}_j)>$ durch eine „Kern-Funktion" $k(\mathbf{x}_i,\mathbf{x}_j)$ als eine implizite Abbildung in den Merkmalsraum, ohne die Anzahl der Parameter zu verändern
- Die Berechnungskomplexitätsordnung wächst nicht, da keine expliziten Abbildungen φ berechnet werden müssen

Kernel-Trick

- Um im Merkmalsraum F zu lernen, ist keine Kenntnis des Merkmalsraumes und keine Kenntnis der zugrunde liegenden Abbildung notwendig!

$$\mathbf{x}=(x_1,\ldots,x_n)\mapsto\varphi(\mathbf{x})=(\varphi_1(\mathbf{x}),\ldots,\varphi_d(\mathbf{x}))$$

$$X\to F=\{\varphi(\mathbf{x})\mid\mathbf{x}\in X\}$$

Kernel-Trick

- Um im Merkmalsraum F zu lernen, ist keine Kenntnis des Merkmalsraumes und keine Kenntnis der zugrunde liegenden Abbildung notwendig!

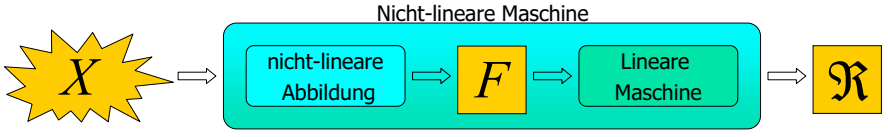
$$\mathbf{x} = (x_1,\ldots,x_n) \mapsto \boldsymbol{\varphi}(\mathbf{x}) = (\varphi_1(\mathbf{x}),\ldots,\varphi_d(\mathbf{x}))$$

$$X \to F = \{\boldsymbol{\varphi}(\mathbf{x}) \mid \mathbf{x} \in X\}$$

- Man kann also in einen unendlich dimensionalen Merkmalsraum effektiv lernen!

Kernel-Trick

- Das Lernen/Trainieren mit einer nicht-linearen Maschine im Eingaberaum ist äquivalent zu einer (festen) nicht-linearen Abbildung des Eingaberaumes in den Merkmalsraum und der Anwendung einer linearen Maschine auf dem Merkmalsraum



Kernel-Trick

- Zusammenfassung des Kernel-Tricks:

$$f(\mathbf{x}) = \sum_{i=1}^{N} \omega_i \cdot \varphi_i(\mathbf{x})$$

$$= \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot <\boldsymbol{\varphi}(\mathbf{x_i}) \cdot \boldsymbol{\varphi}(\mathbf{x})> \quad \text{duale Darstellung}$$

$$= \sum_{i=1}^{l} \alpha_i \cdot y_i \cdot k(\mathbf{x_i}, \mathbf{x}) \quad \text{Kernel-Maschine}$$

Anforderungen an Kerne

Anforderungen an Kerne

Definition:

Ein Kern ist eine Funktion k, so daß für alle $\mathbf{x}, \mathbf{y} \in X$ gilt:

$$k(\mathbf{x},\mathbf{y}) = <\varphi(\mathbf{x}) \cdot \varphi(\mathbf{y})>$$

Hierbei ist $\varphi : X \rightarrow F$ eine Abbildung in den Merkmalsraum.

Anforderungen an Kerne

Notwendige Bedingungen:

- Symmetrie

$$k(\mathbf{x},\mathbf{z}) = <\varphi(\mathbf{x}) \cdot \varphi(\mathbf{z})> = <\varphi(\mathbf{z}) \cdot \varphi(\mathbf{x})> = k(\mathbf{z},\mathbf{x})$$

- Erfüllung der Cauchy-Schwartz-Ungleichung

$$k(\mathbf{x},\mathbf{z})^2 = <\varphi(\mathbf{x}) \cdot \varphi(\mathbf{z})>^2$$

$$\leq \|\varphi(\mathbf{x})\|^2 \cdot \|\varphi(\mathbf{z})\|^2 = <\varphi(\mathbf{x}) \cdot \varphi(\mathbf{x})> \cdot <\varphi(\mathbf{z}) \cdot \varphi(\mathbf{z})>$$

$$= k(\mathbf{x},\mathbf{x}) \cdot k(\mathbf{z},\mathbf{z})$$

Anforderungen an Kerne

Notwendige Bedingungen:

- Symmetrie

$$k(\mathbf{x},\mathbf{z}) = <\boldsymbol{\varphi}(\mathbf{x})\cdot\boldsymbol{\varphi}(\mathbf{z})> = <\boldsymbol{\varphi}(\mathbf{z})\cdot\boldsymbol{\varphi}(\mathbf{x})> = k(\mathbf{z},\mathbf{x})$$

- Erfüllung der Cauchy-Schwartz-Ungleichung

$$k(\mathbf{x},\mathbf{z})^2 = <\boldsymbol{\varphi}(\mathbf{x})\cdot\boldsymbol{\varphi}(\mathbf{z})>^2$$

$$\leq \left\|\boldsymbol{\varphi}(\mathbf{x})\right\|^2 \cdot \left\|\boldsymbol{\varphi}(\mathbf{z})\right\|^2 = <\boldsymbol{\varphi}(\mathbf{x})\cdot\boldsymbol{\varphi}(\mathbf{x})> \cdot <\boldsymbol{\varphi}(\mathbf{z})\cdot\boldsymbol{\varphi}(\mathbf{z})>$$

$$= k(\mathbf{x},\mathbf{x})\cdot k(\mathbf{z},\mathbf{z})$$

Nicht hinreichend!

Anforderungen an Kerne

positiv semi-definite Kern-Matrizen

- alle Eigenwerte sind nicht-negativ
- Sei $\mathbf{K}$ eine beliebige symmetrische positiv semi-definite Matrix. Dann gilt $\forall \mathbf{x} : \mathbf{x}^t \cdot \mathbf{K} \cdot \mathbf{x} \geq 0$

Anforderungen an Kerne

positiv semi-definite Kern-Matrizen

- alle Eigenwerte sind nicht-negativ
- Sei $\mathbf{K}$ eine beliebige symmetrische positiv semi-definite Matrix. Dann gilt $\forall \mathbf{x} : \mathbf{x}^t \cdot \mathbf{K} \cdot \mathbf{x} \geq 0$

Jede positiv semi-definite symmetrische Matrix kann als eine Kern-Matrix betrachtet werden und umgekehrt ist jede Kern-Matrix symmetrisch positiv semi-definit.

Bei einer Kern-Funktion sind alle entsprechenden Gram-Matrizen symmetrisch positiv semi-definit.

Anforderungen an Kerne

Beweis, daß eine symmetrische positiv-definite Matrix den Anforderungen einer Kern-Matrix genügt.

Sei $\mathbf{K}$ also eine symmetrische positiv semi-definite Matrix. Wegen der geforderten Symmetrie kann $\mathbf{K}$ diagonalisiert werden, also gibt es eine Matrix $\mathbf{V}$, so daß gilt: $\mathbf{K} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^t$. $\mathbf{\Lambda}$ ist eine Diagonalmatrix mit den Eigenwerten λ_i von $\mathbf{K}$ und den jeweiligen Eigenvektoren $\mathbf{v}_i$ als Spalten von $\mathbf{V}$.

Man definiert nun folgende Abbildung: $\varphi : \mathbf{x}_i \mapsto \left(\sqrt{\lambda_t} \cdot v_{ti}\right)_{t=1}^{n}$

Anforderungen an Kerne

Mit dieser Abbildung gilt nun für das Skalarprodukt:

$$< \varphi(\mathbf{x_i}) \cdot \varphi(\mathbf{x_j}) > = \sum_{t=1}^{n} \lambda_t \cdot v_{ti} \cdot v_{tj} = \left(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^t\right)_{ij} = \mathbf{K}_{ij} = k(\mathbf{x_i}, \mathbf{x_j})$$

$\mathbf{K}_{ij}$ entspricht den Einträgen der Gram-Matrix, so daß $k(\mathbf{x_i}, \mathbf{x_j})$ offensichtlich eine Kernfunktion ist. □

Umgekehrt gilt bei einer Kern-Matrix wegen der Kommutativität des Skalarproduktes die Symmetrie. Die positiv semi-Definitheit der Kernmatrix ergibt sich aus den Eigenschaften des zugrundeliegenden Hilbertraumes, der u.a. normiert ist und deshalb keine negativen Werte des Skalarproduktes zuläßt.

Anforderungen an Kerne

Mercer-Theorem:

„konstruiere eine Abbildung φ implizit durch einen Kern k, d.h., eine Abbildung φ, so daß k das Skalarprodukt in denjenigen Raum berechnet, auf den φ abbildet"

Sei X kompakt, also beschränkt und abgeschlossen* und weiterhin sei $k : X \times X \rightarrow \Re$ eine Abbildung, wobei hier gilt: $k \in \underbrace{\left(L_2(X \times X)\right) \cap \left(C^0(X \times X)\right)}$

Einfach gesagt: die Abbildung k muß stetig und das Quadrat von k muß integrierbar sein

Anforderungen an Kerne

Sofern nun $\forall f \in L_2(X): \iint_{X\times X} k(\mathbf{x},\mathbf{x}') \cdot f(\mathbf{x}) \cdot f(\mathbf{x}')\, d\mathbf{x} d\mathbf{x}' \geq 0$ gilt, gibt es eine Zerlegung von $k(\mathbf{x},\mathbf{x}')$ in ein Eigensystem mit positiven Eigenwerten λ_i und Eigenfunktionen $\phi_i(\cdot)$, d.h.

$$k(\mathbf{x},\mathbf{x}') = \sum_i \lambda_i \cdot \varphi_i(\mathbf{x}) \cdot \varphi_i(\mathbf{x}')$$

Das Mercer-Theorem gibt also notwendige und hinreichende Bedingungen für eine stetige Funktion $k(\mathbf{x},\mathbf{x}')$ um obige Repräsentation zu garantieren.

Konstruktionsregeln

Konstruktionsregeln

Seien folgende Objekte gegeben:

- k_1 und k_2 Kerne über $X \times X$ wobei $X \subseteq \Re^n$
- $a \in \Re^+$ eine reelle positive Zahl
- $f(\cdot)$ sei eine reelle Funktion
- φ sei eine Abbildung mit einem Kern k_3 über $\Re^m \times \Re^m$
- $\mathbf{B}$ eine symmetrische positiv semi-definite $n \times n$ Matrix

Konstruktionsregeln

- Dann sind die folgenden Funktionen Kerne:
  1. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$

Konstruktionsregeln

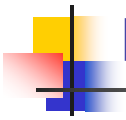
- Dann sind die folgenden Funktionen Kerne:
  1. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$
  2. : $k(\mathbf{x},\mathbf{z}) = a \cdot k_1(\mathbf{x},\mathbf{z})$

Konstruktionsregeln

- Dann sind die folgenden Funktionen Kerne:
  1. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$
  2. : $k(\mathbf{x},\mathbf{z}) = a \cdot k_1(\mathbf{x},\mathbf{z})$
  3. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) \cdot k_2(\mathbf{x},\mathbf{z})$

Konstruktionsregeln

- Dann sind die folgenden Funktionen Kerne:
  1. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$
  2. : $k(\mathbf{x},\mathbf{z}) = a \cdot k_1(\mathbf{x},\mathbf{z})$
  3. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) \cdot k_2(\mathbf{x},\mathbf{z})$
  4. : $k(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) \cdot f(\mathbf{z})$



Konstruktionsregeln

- Dann sind die folgenden Funktionen Kerne:
  1. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$
  2. : $k(\mathbf{x},\mathbf{z}) = a \cdot k_1(\mathbf{x},\mathbf{z})$
  3. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) \cdot k_2(\mathbf{x},\mathbf{z})$
  4. : $k(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) \cdot f(\mathbf{z})$
  5. : $k(\mathbf{x},\mathbf{z}) = k_3(\boldsymbol{\varphi}(\mathbf{x}),\boldsymbol{\varphi}(\mathbf{z}))$

Konstruktionsregeln

- Dann sind die folgenden Funktionen Kerne:

1. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$
2. : $k(\mathbf{x},\mathbf{z}) = a \cdot k_1(\mathbf{x},\mathbf{z})$
3. : $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) \cdot k_2(\mathbf{x},\mathbf{z})$
4. : $k(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) \cdot f(\mathbf{z})$
5. : $k(\mathbf{x},\mathbf{z}) = k_3(\boldsymbol{\varphi}(\mathbf{x}),\boldsymbol{\varphi}(\mathbf{z}))$
6. : $k(\mathbf{x},\mathbf{z}) = \mathbf{x}' \cdot \mathbf{B} \cdot \mathbf{z}$

Konstruktionsregeln (Beweise)

1. $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) + k_2(\mathbf{x},\mathbf{z})$

Sei $\boldsymbol{\alpha} \in \Re^l$. Dann gilt:

$$\boldsymbol{\alpha}'(\mathbf{K}_1 + \mathbf{K}_2)\boldsymbol{\alpha} = \underbrace{\boldsymbol{\alpha}'\mathbf{K}_1\boldsymbol{\alpha}}_{\mathbf{K}_1 \text{ symm.pos.def.}} + \underbrace{\boldsymbol{\alpha}'\mathbf{K}_2\boldsymbol{\alpha}}_{\mathbf{K}_2 \text{ symm.pos.def.}} \geq 0$$

Konstruktionsregeln (Beweise)

2. $k(\mathbf{x},\mathbf{z}) = a \cdot k_1(\mathbf{x},\mathbf{z})$

Sei $\boldsymbol{\alpha} \in \Re^l$. Dann gilt:

$$\boldsymbol{\alpha}' \cdot a \cdot \mathbf{K}_1 \cdot \boldsymbol{\alpha} = a \cdot \underbrace{\boldsymbol{\alpha}' \cdot \mathbf{K}_1 \cdot \boldsymbol{\alpha}}_{\mathbf{K}_1 \text{ symm. pos. def.}} \geq 0$$

Konstruktionsregeln (Beweise)

3. $k(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z}) \cdot k_2(\mathbf{x},\mathbf{z})$

Sei $\mathbf{K} = \mathbf{K}_1 \otimes \mathbf{K}_2$ das Tensorprodukt von $\mathbf{K}_1$ und $\mathbf{K}_2$

Das Tensorprodukt zweier positiv semi-definiter Matrizen ist wiederum positiv semi-definit, da die Eigenwerte des Produktes alle Paare von Produkten der Eigenwerte der beiden Komponenten sind.

Die Matrix entsprechend der Funktion $k_1 k_2$ ergibt sich aus der komponentenweisen Multiplikation der Elemente der beiden Matrizen (auch als „Schur-Matrix" $\mathbf{H}$ bekannt)
$\mathbf{H}$ ist eine Untermatrix von $\mathbf{K}$, so daß für alle $\boldsymbol{\alpha} \in \Re^l$ ein $\boldsymbol{\alpha}_1 \in \Re^{l^2}$ existiert, wobei hier dann gilt: $\boldsymbol{\alpha}' \mathbf{H} \boldsymbol{\alpha} = \boldsymbol{\alpha}_1' \mathbf{K} \boldsymbol{\alpha}_1 \geq 0$

Konstruktionsregeln (Beweise)

4. $k(\mathbf{x},\mathbf{z}) = f(\mathbf{x}) \cdot f(\mathbf{z})$

Es gilt offensichtlich für die Bilinearformen:

$$\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j k(\mathbf{x_i},\mathbf{x_j}) = \sum_i\sum_j \alpha_i\alpha_j f(\mathbf{x_i}) f(\mathbf{x_j}) =$$

$$= \sum_i \alpha_i f(\mathbf{x_i}) \sum_j \alpha_j f(\mathbf{x_j}) = \left(\sum_i \alpha_i f(\mathbf{x_i})\right)^2 \geq 0$$

Konstruktionsregeln (Beweise)

5. $k(\mathbf{x},\mathbf{z}) = k_3(\boldsymbol{\varphi}(\mathbf{x}),\boldsymbol{\varphi}(\mathbf{z}))$

Da nach Voraussetzung k_3 eine Kernfunktion ist,

also positiv semi-definit, ist auch seine

Einschränkung auf die Punkte $\boldsymbol{\varphi}(\mathbf{x_1})\ldots\boldsymbol{\varphi}(\mathbf{x_l})$

positiv semi-definit.

Konstruktionsregeln (Beweise)

6. $k(\mathbf{x},\mathbf{z}) = \mathbf{x}'\mathbf{B}\mathbf{z}$

Sei $\mathbf{B} = \mathbf{V}'\cdot\Lambda\cdot\mathbf{V}$ und $\mathbf{A} = \sqrt{\Lambda}\cdot\mathbf{V}$

Es gilt: $k(\mathbf{x},\mathbf{z}) = \mathbf{x}'\cdot\mathbf{B}\cdot\mathbf{z} = \mathbf{x}'\cdot\mathbf{V}'\cdot\Lambda\cdot\mathbf{V}\cdot\mathbf{z} =$

$= \mathbf{x}'\cdot\mathbf{V}'\cdot\sqrt{\Lambda}\cdot\sqrt{\Lambda}\cdot\mathbf{V}\cdot\mathbf{z} = \mathbf{x}'\cdot\mathbf{A}'\cdot\mathbf{A}\cdot z =$

$= <\mathbf{A}\mathbf{x}\cdot\mathbf{A}\mathbf{z}>$

Bemerkung: $\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$ ist eine beliebige (feste) lineare Transformation für eine Matrix $\mathbf{A}$

□

Problemspezifische Kerne

Problemspezifische Kerne

- Beschränkung auf Vektorraumdaten, insbesondere Muster- / Bilderkennung
- Anwendung auf Nichtvektorraumdaten, beispielsweise Texterkennung, im Vortrag von Jingtao Wang nächste Woche

Problemspezifische Kerne (Bild- / Mustererkennung)

- „traditionelle" Methoden sind schlecht wegen der hohen Dimension der Eingabedaten:
  Dim = Breite*Höhe (*3)
- Anwendung der SVM beispielsweise auf Histogrammdaten (globale, invariante Bilddaten) ⇒ „Abstand" zweier Histogramme
  (durch Abstandsmessung der Histogrammvektoren mit einem geeigneten Abstandsmaß)

Problemspezifische Kerne (Bild- / Mustererkennung)

- Ausnutzung der lokalen Korrelation von Farbwerten innerhalb der Bilddaten
- Einfache und „problemunspezifische" Kerne liefern bemerkenswerte Resultate
  (insbesondere im Vergleich zu hoch-angepaßten KNNs)
- Bekanntes Vorwissen kann bei der Wahl des Kernes und seiner Parameterbelegung einfließen

Problemspezifische Kerne (Zeichenerkennung)

- Einfache Ansätze:
  - Mögliche („unspezifische") allgemeine Kerne für Zeichenerkennung:

$$k(\mathbf{x},\mathbf{y}) = \left(\frac{<\mathbf{x}\cdot\mathbf{y}>}{n^2} \right)^d$$

$$k(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{n^2 \cdot \sigma^2} \right)$$

Eingabedimension: *(n,n)*
Grad der Polynome: 1 - 6
Varianz: 0.1 - 4.0

Problemspezifische Kerne (Zeichenerkennung)

- Bemerkungen:
  - Trotz des allgemeinen Ansatzes erzielen die SVMs im Schnitt bessere Ergebnisse als beispielsweise hoch angepasste Neuronale Netze

Problemspezifische Kerne (Zeichenerkennung)

- Bemerkungen:
  - Trotz des allgemeinen Ansatzes erzielen die SVMs im Schnitt bessere Ergebnisse als beispielsweise hoch angepasste Neuronale Netze
  - insbesondere RBF-Kerne erlauben durch ihre Struktur die Berücksichtigung von *a-priori* - Wahrscheinlichkeiten (Streuung der Daten als Maß für σ)

Problemspezifische Kerne (Zeichenerkennung)

- Bemerkungen:
  - Trotz des allgemeinen Ansatzes erzielen die SVMs im Schnitt bessere Ergebnisse als beispielsweise hoch angepasste Neuronale Netze
  - insbesondere RBF-Kerne erlauben durch ihre Struktur die Berücksichtigung von *a-priori* - Wahrscheinlichkeiten (Streuung der Daten als Maß für σ)
  - Aus dem allgemeinen Ansatz entsteht ein problemspezifisches Vorgehen…

Problemspezifische Kerne (Bild- / Mustererkennung)

- Konkrete Vorgehensweise:
  - Finde eine geeignete Abbildung φ vom Bildraum in den Merkmalsraum (beispielsweise Histogramm als rotations-, translations- und skalierungsinvariante Abbildung)
  - Finde ein Kriterium, um den Abstand zweier Merkmalsvektoren (geeignet) zu messen
  - Allgemeiner Ansatz: $k(\mathbf{x},\mathbf{z}) = \exp\left(-\frac{d(\mathbf{x},\mathbf{z})}{\sigma^2}\right)$

Problemspezifische Kerne (Bild- / Mustererkennung)

- Vorgehensweise:
  - Mögliche Ansätze für Abstandsmessung:

$$d(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{n} \frac{(x_i - z_i)^2}{x_i + z_i} \qquad \chi^2 \; Funktion$$

$$d_p(\mathbf{x},\mathbf{z}) = \left(\sum_{i=1}^{n} \left| x_i - z_i \right|^p \right)^{\frac{1}{p}} \qquad p - Norm$$

p=1,2 erfüllt das Mercer-Theorem

Zusammenfassung

- Komplexe problemspezifische Kerne können aus einfachen, weitgehend problem-unspezifischen Kernen aufgebaut werden
- Voraussetzungen: Mercer-Theorem, u.a.
- Besonderheiten in der Problemstellung, (z.B.: lokale Korrelation bei Bildern), sollten beim Kerndesign berücksichtigt werden (Lokalkorrelationskerne)

Literatur

- Kernel-Trick

  C.J.C. Burges. A tutorial on SVMs for pattern recogn. *Data Mining and Knowledge Discovery*, 1998

- Kern-Theorie

  N. Cristianini, J. Shawe-Taylor. Introduction to SVMs. C*ambridge University Press*, 2000

  B. Schölkopf, A. Smola. Learning with kernels. *MIT-Press*, 2002

Erläuterungen:

- Mercer-Theorem*:

  ** als Eigenschaften mathematischer Objekte

  • beschränkt** : $\exists c \in \Re^+ : \forall \mathbf{x} \in X : |f(\mathbf{x})| \leq c$

  • abgeschlossen**: jede Cauchy-Folge konvergiert:

  $$\forall n, m > N \;\; \exists \varepsilon > 0 : \|x_n - x_m\| < \varepsilon$$

  • der Hilbert-Raum $L_2(X)$ hat folgende Eigenschaften:

  $$\|f(x)\|_{L_2} = \int_X f(x)^2 dx < \infty \text{ und } < f \cdot g >= \int_X f(x) \cdot g(x) dx$$

  • die Menge $C^0(X)$ ist Menge der stetigen Funktionen (deren Quadrat integrierbar ist)

Notation

N	Dimension des Merkmalsraumes
$Y(\Re)$	Ausgaberaum
X	Eingaberaum
F	Merkmalsraum
$< x \cdot y >$	Skalarprodukt
$\varphi : X \rightarrow F$	Abbildung des Ursprungsraumes in Merkmalsraum
$k(\mathbf{x}, \mathbf{y})$	Kernfunktion
$\mathbf{K}$	Matrix (Kern-Matrix)
$\boldsymbol{\omega}$	Gewichtsvektor
α	Lagrange-Multiplikatoren
$\Re$	reelle Zahlen
l	Anzahl der Trainingsbeispiele