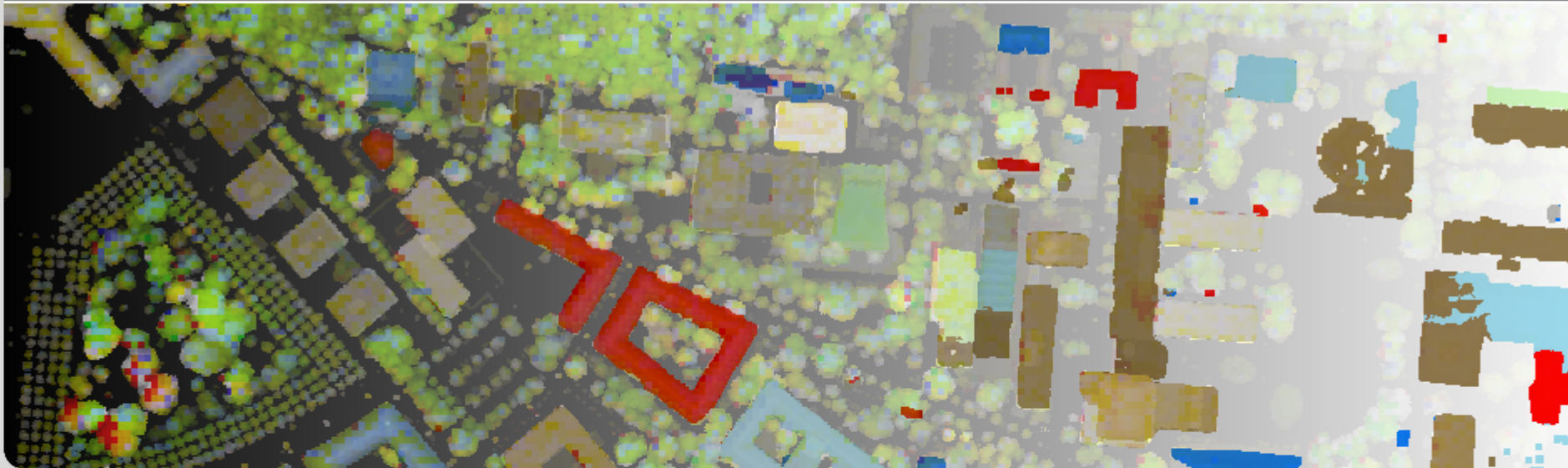


Kernel-Composition als Methode der Datenfusion in der SVM Klassifizierung

DGPF Workshop „Neue Methoden für die Auswertung und Interpretation von Fernerkundungsdaten“
Andreas Ch. Braun

Institute of Photogrammetry and Remote Sensing - IPF



Inhalt und Gliederung

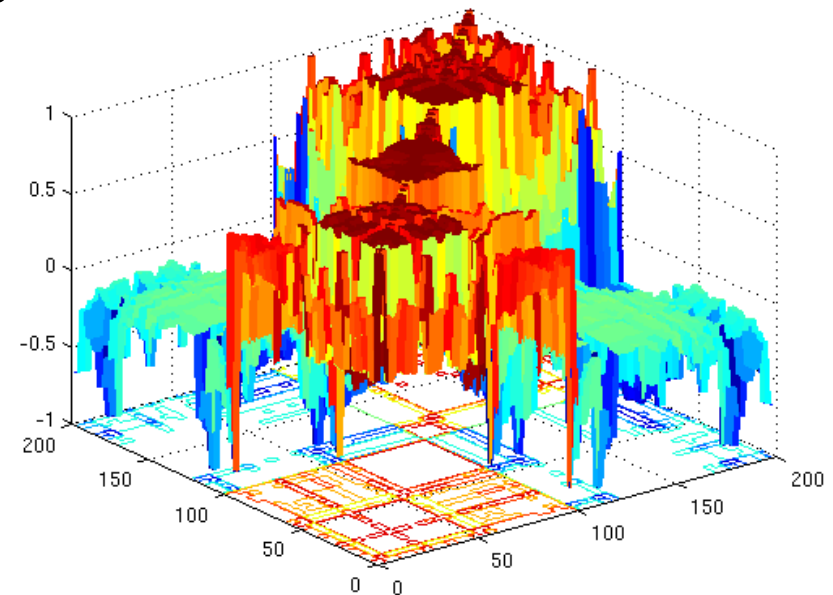
- Einführung Kernel Matrizen und Kernel Composition

- Beispiel 1: Datenfusion
 - Methodik
 - Anwendung

- Beispiel 2: Semi-Supervised Learning
 - Methodik
 - Anwendung

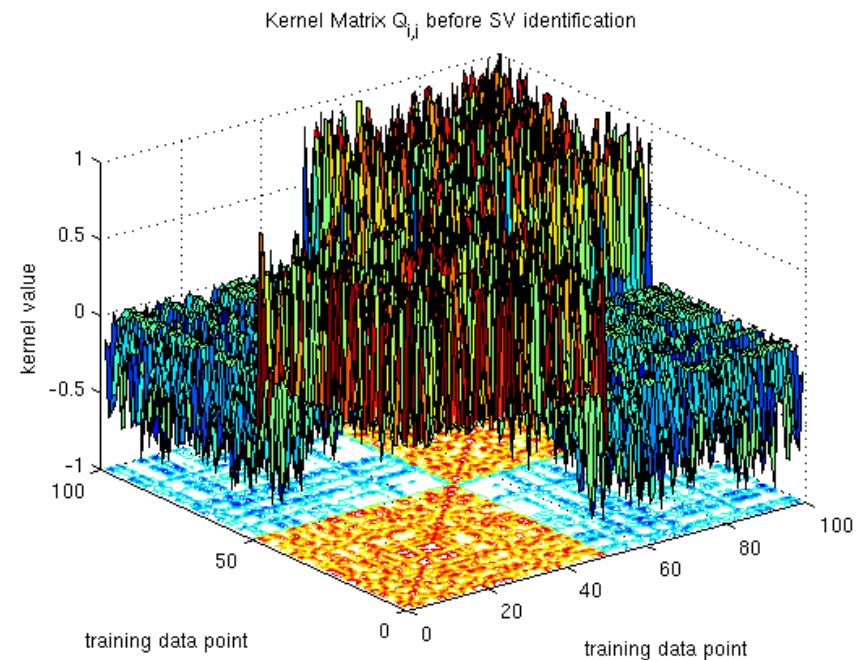
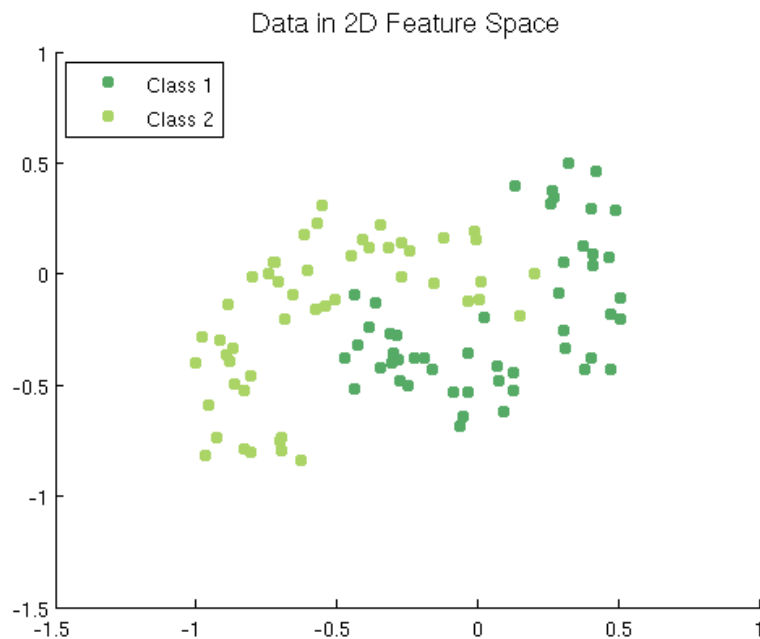
- Beispiel 3: Change Detection
 - Methodik
 - Anwendung

- Kurzer Literaturüberblick & Fazit



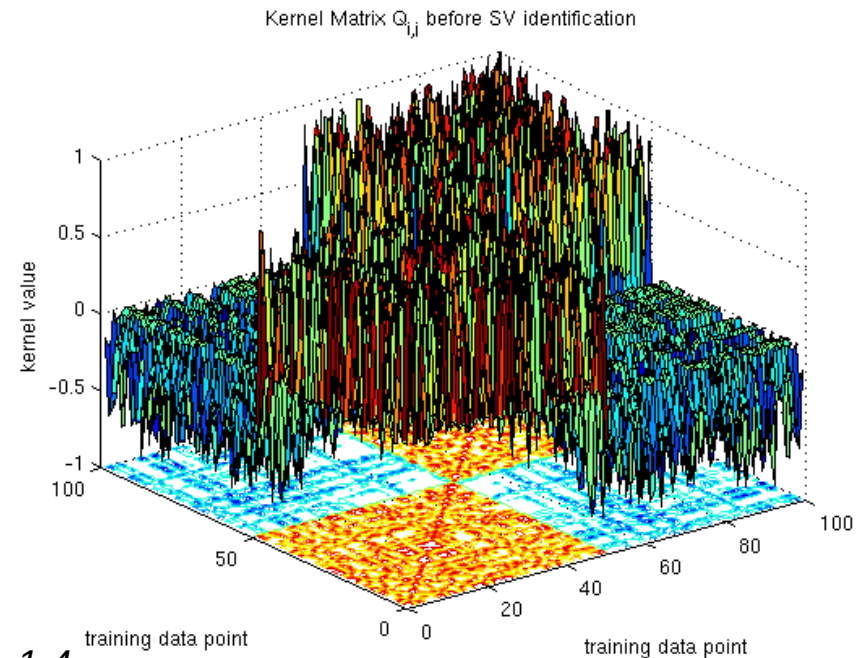
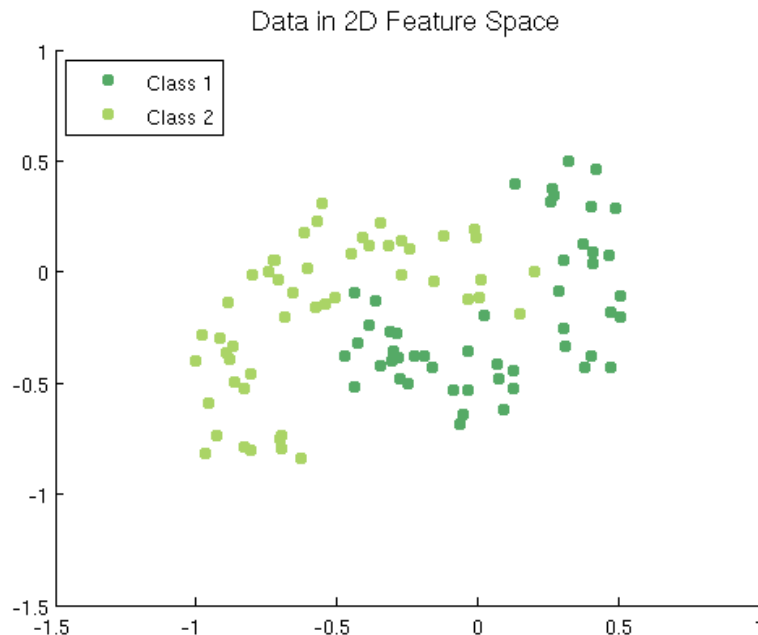
Kernel Matrizen

- Kernel Matrizen sind Distanzmatrizen
 - RBF: $K(x_i, x_j) = \exp(-\sigma \|x_i - x_j\|^2)$
 - Euklidische Distanz



Kernel Matrizen

- Kernel Matrizen sind Distanzmatrizen
 - $\theta(x_i - x_j) = \arccos(x_i \cdot x_j / \|x_i\| \|x_j\|)$ spektraler Winkel
 - RBF: $K(x_i, x_j) = \exp(-\sigma \theta(x_i - x_j))$
 - Honeine, Richard, 2010



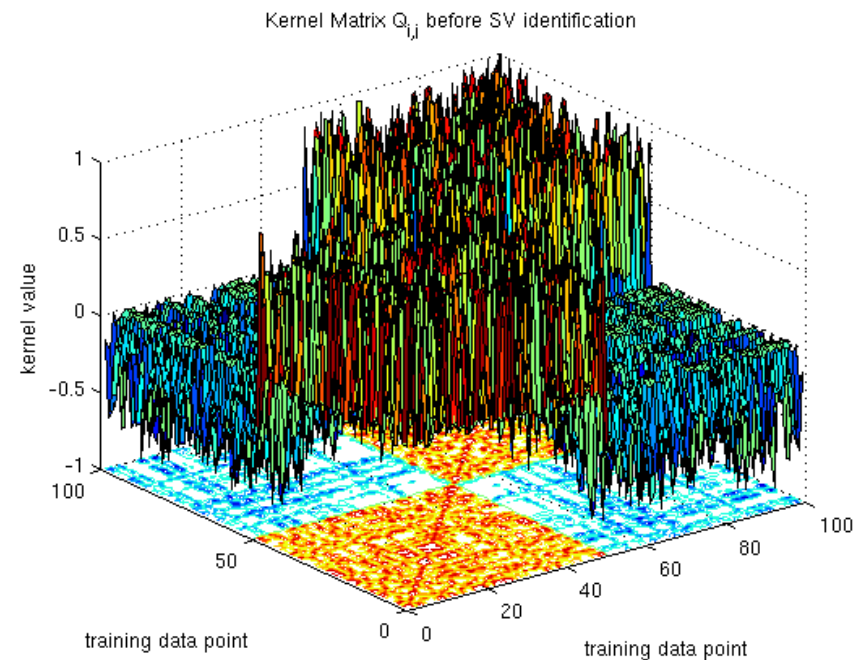
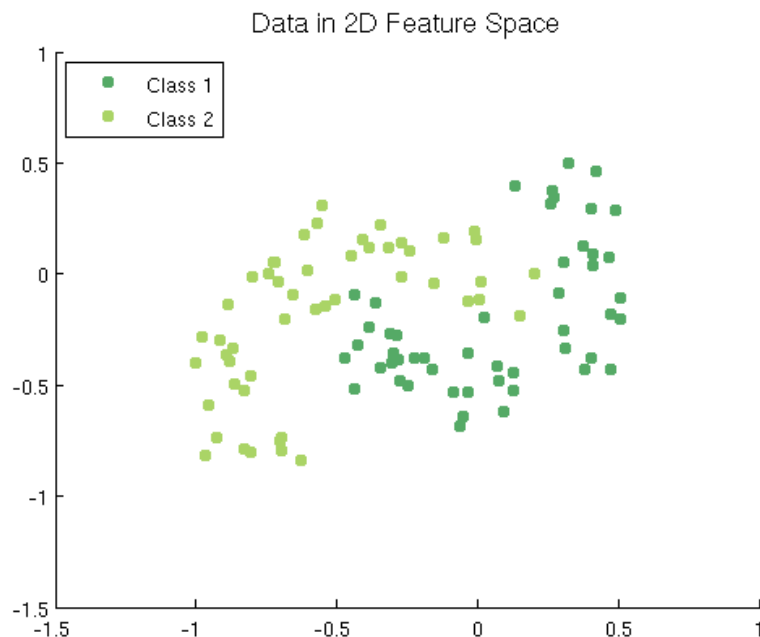
Honeine, P. & Richard, C. In Proc. WHISPERS, 2010, pp. 1-4

Kernel Matrizen

■ SVM Problem in der dualen Formulierung

maximiere: $L_D(w, b, \Lambda) = \sum \lambda_i - \frac{1}{2} \sum \sum \lambda_i \lambda_j y_i y_j K(x_i, x_j)$

NB: $\lambda_i > 0 \forall$ Support Vectors

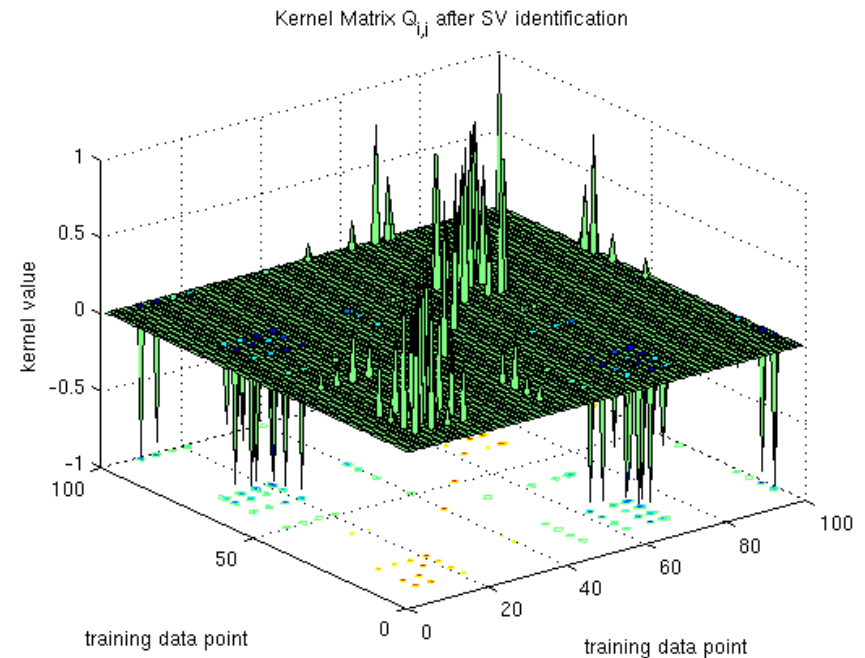
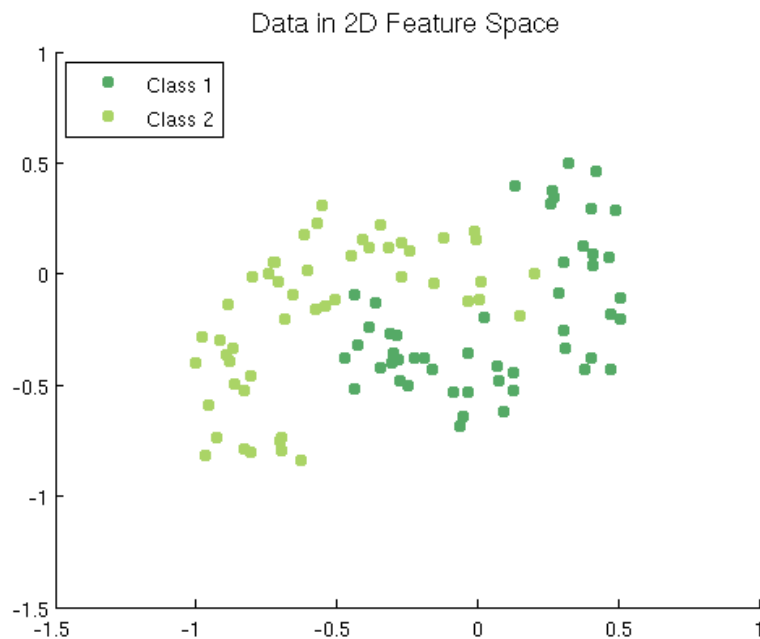


Kernel Matrizen

■ SVM Problem in der dualen Formulierung

maximiere: $L_D(w, b, \Lambda) = \sum \lambda_i - \frac{1}{2} \sum \sum \lambda_i \lambda_j y_i y_j K(x_i, x_j)$

NB: $\lambda_i > 0 \forall$ Support Vectors

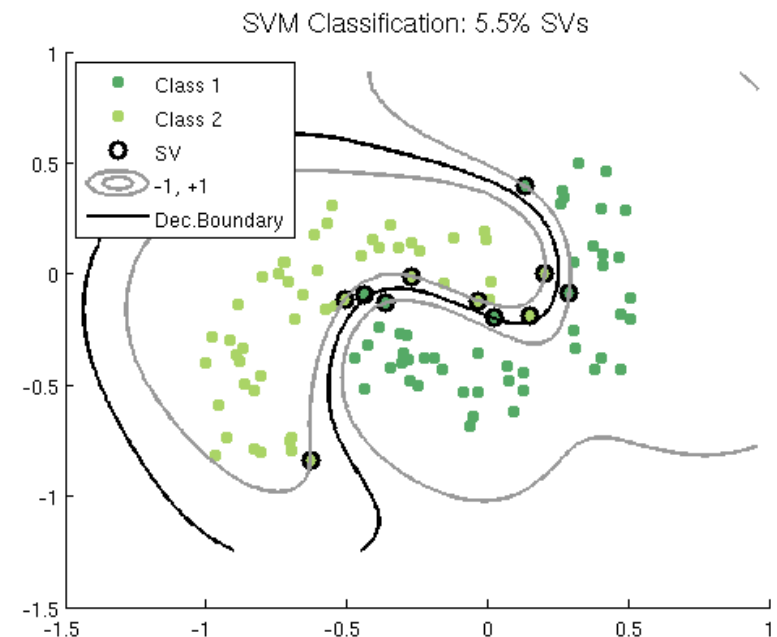
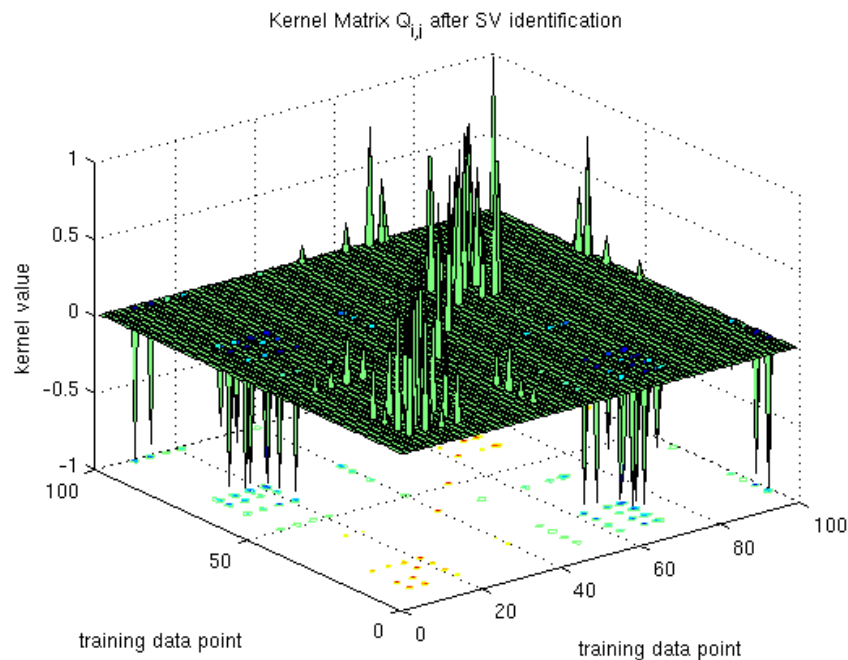


Kernel Matrizen

■ SVM Problem in der dualen Formulierung

maximiere: $L_D(w, b, \Lambda) = \sum \lambda_i - \frac{1}{2} \sum \sum \lambda_i \lambda_j y_i y_j K(x_i, x_j)$

NB: $\lambda_i > 0 \forall$ Support Vectors



Kernel Composition Datenfusion

■ Theorem von Mercer: verschiedene Kernelfunktionen $K(x_i, x_j)$ können verknüpft werden:

■ Direct Summation Kernel:

$$K_C = K_A(x_i(A), x_j(A)) + K_B(x_i(B), x_j(B))$$

■ Weighted Summation Kernel:

$$K_C = f_1 * K_A(x_i(A), x_j(A)) + f_2 * K_B(x_i(B), x_j(B)), \quad \text{mit } f_2 = 1 - f_1$$

■ Product Kernel:

$$K_C = K_A(x_i(A), x_j(A)) * K_B(x_i(B), x_j(B))$$

■ Cross-Information Kernel:

$$K_C = K_A(x_i(A), x_j(A)) + K_B(x_i(B), x_j(B)) + K_{AB}(x_i(A), x_j(B)) + K_{BA}(x_i(B), x_j(A))$$

Kernel Composition Datenfusion

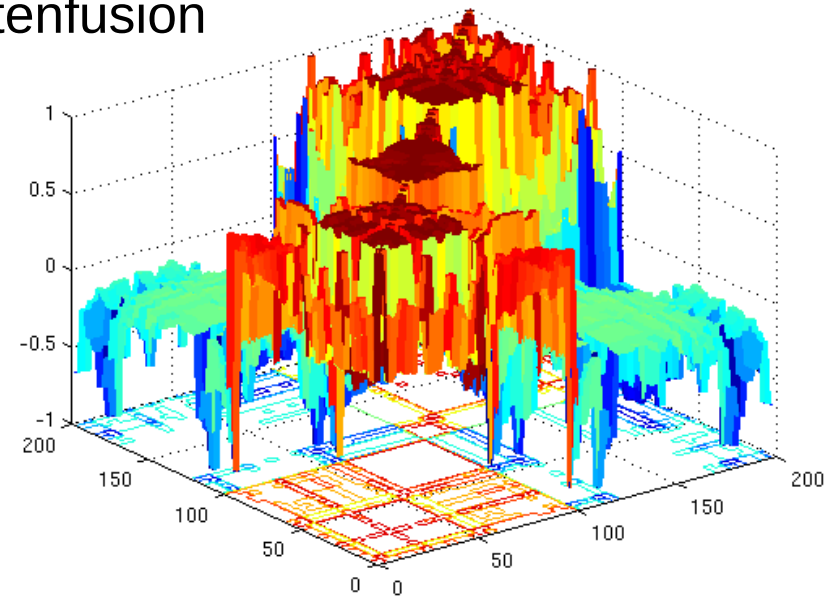
- Camps-Valls et al. (2005) nutzen diese Idee erstmals um Daten zu fusionieren

- AVIRIS Indian Pines und Bildfilter ($[\mu]$, $[\mu, \sigma]$ in 3x3, 9x9 Fenster)
 - Gesamtgenauigkeit: Stacked Features Fusion: **94,2%**
 - Gesamtgenauigkeit: Kernel Composition: **96,6%**

Camps-Valls, G. & Gomez-Chova, L. & Vila-Frances, J. & Calpe-Maravilla, J., GRSI 3(1), 2005, pp. 93-97

Anwendungsbeispiel 1

Kernel Composition zur Datenfusion

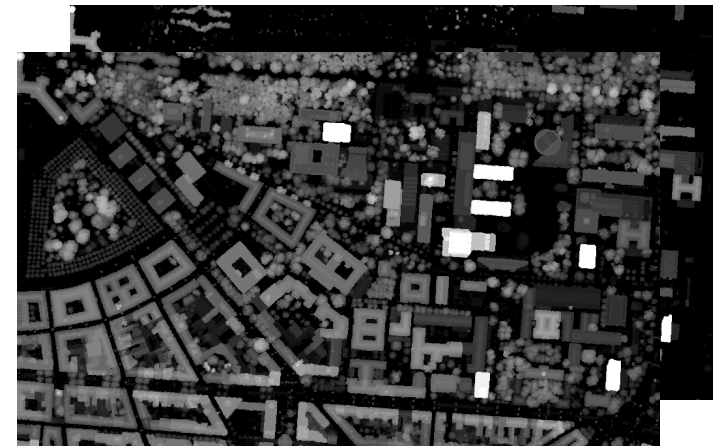
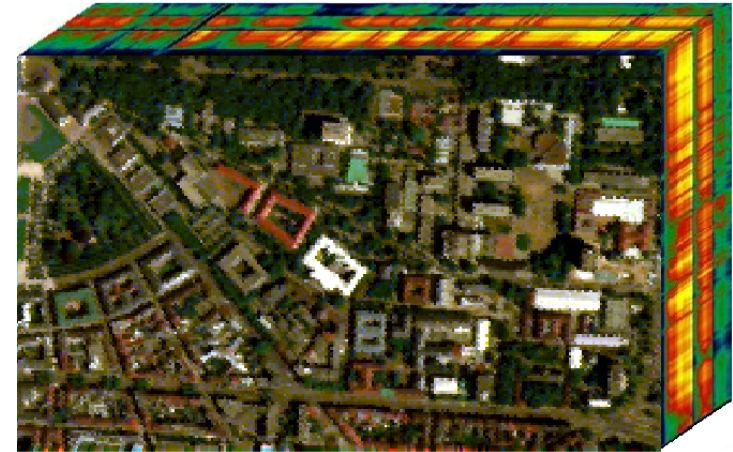


Kernel Composition Datenfusion

- Sehr simpler, flexibler Ansatz
 - Berechne K_1 auf Datenquelle 1
 - Berechne K_2 auf Datenquelle 2
 - Fusioniere, z.b. $K_n = K_1 + K_2$
 - SVM lernt dann K_n
- **Vorteil** (speziell hyperspektral). Direktes Aneinanderhängen von Merkmalsvektoren erhöht die Dimensionalität. Hughes Phänomen entsteht (Curse of Dimensionality).
Kernel Composition arbeitet im hoch-dimensionalen Raum der SVM (RKHS) → Hughes Phänomen wird umgangen.

Beispiel 1: Fusion Hyperspektral und Laserscanning

- Ziel: Schadstoffströme von Dächern abschätzen
Braun et al. 2011
- Dachmaterialien identifizieren
 - Verschiedene spektrale Eigenschaften
 - Verschiedene Dachformen
- Zwei Datensätze fusionieren
 - Airborne laser scanning
 - Hyperspektral (HyMap)
- Dann SVM Klassifikation

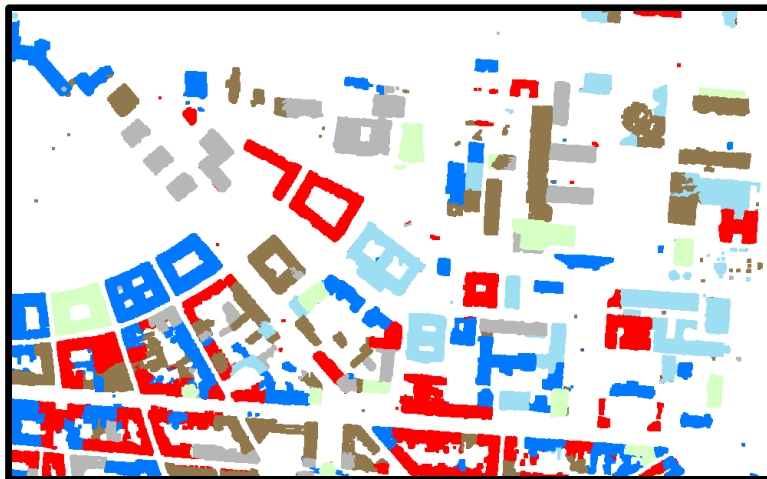


Braun, A.C. & Weidner, U. & Jutzi, B. & Hinz, S. In Proc. ISPRS Hannover Workshop 2011

Beispiel 1: Fusion Hyperspektral und Laserscanning

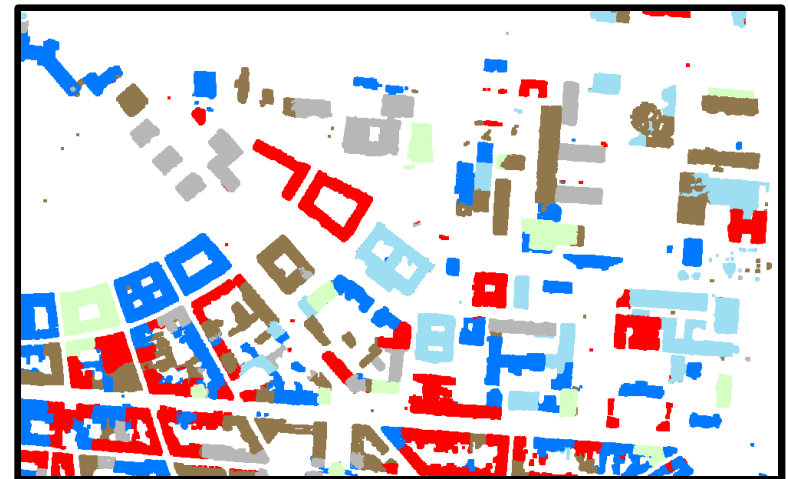
■ Daten Fusion: Concatenation

- OAA: 83.9%
- Kappa: 0.79
- min(Cp.): 75.3%
- min(Cr.): 58.9%
- Dauer: 72min



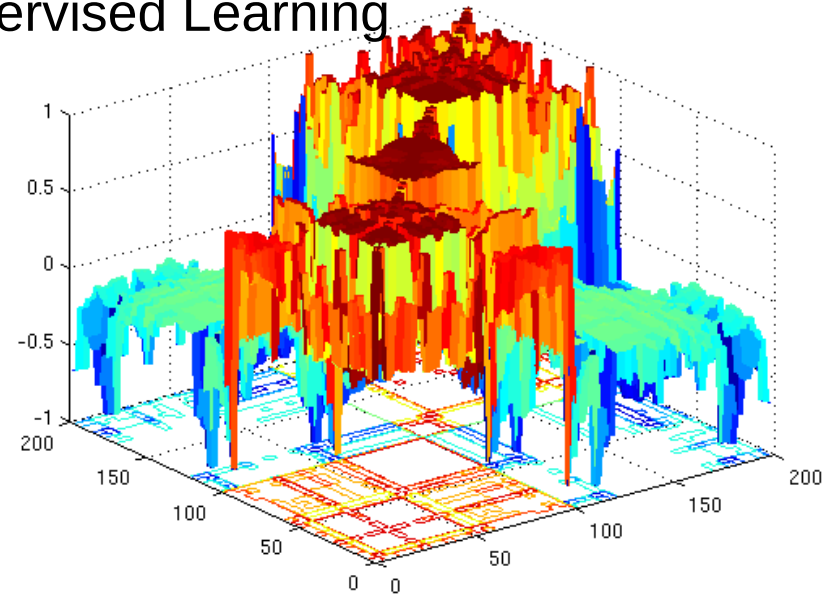
■ Daten Fusion: Kernel Comp.

- Alle Cp., Cr. Werte höher
- OAA: 86.4%
- Kappa: 0.82
- min(Cp.): 80.0%
- min(Cr.): 67.4%
- Dauer: 59min



Anwendungsbeispiel 2

Kernel Composition zum Semi-Supervised Learning



Kernel Composition Semi-Supervised

- Semi-Supervised Learning (SSL). Man will ungelabelte Pixel im Training zusammen mit gelabelten nutzen.
- Konsistenzannahme:
 - Benachbarte Punkte haben sehr wahrscheinlich identische Klassenlabels.
 - Punkte im selben Cluster haben wahrscheinlich identische Klassenlabels.
- Im Prinzip sehr ähnlich: einmal Ortsraum, einmal Merkmalsraum.
- Diese Konsistenzannahme soll ausgenutzt werden, um ungelabelte Pixel im Training zu nutzen.

Kernel Composition Semi-Supervised

- Tuia, Camps-Valls (2009) nutzen hier die Kernel Composition.
 - Erstelle einen SVM Kernel K_{svm} .
 - Berechne z.B. ein K-Means Clustering hinreichend oft.
 - Notiere, welche ungelabelten Pixel mit den Trainingspixeln in den selben Cluster fallen.
 - Berechne einen neuen Kernel K_{bag} : wie oft fallen Punkte ins selbe Cluster.
 - Kernel Composition $K_{neu} = K_{svm} + K_{bag}$: SVM updaten.

Tuia, D., & Camps-Valls, G., In Proc. JURSE 2009, pp. 1-5

Beispiel 2: Semi-Supervised Learning

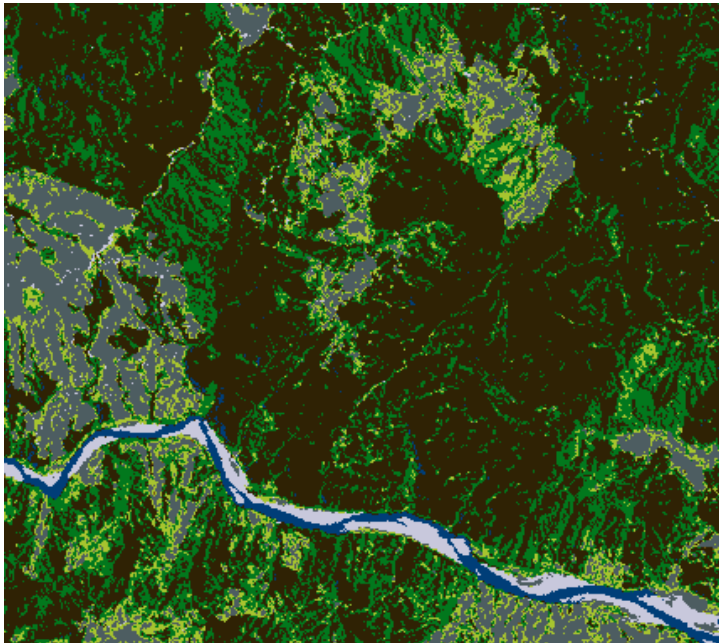
- Ziel: Landnutzungsklassifizierung Chile.
Braun et al. unpublished
- Vegetationseinheiten klassifizieren.
 - Hohe spektrale Ähnlichkeiten.
 - „Semantische“ Klassen (Buschland).
- Landsat Datensätze.
 - Kaum gelabelte Pixel (50 pro Klasse).
 - SSL Kbag nutzen.
- Dann SVM Klassifikation.



Beispiel 2: Semi-Supervised Learning

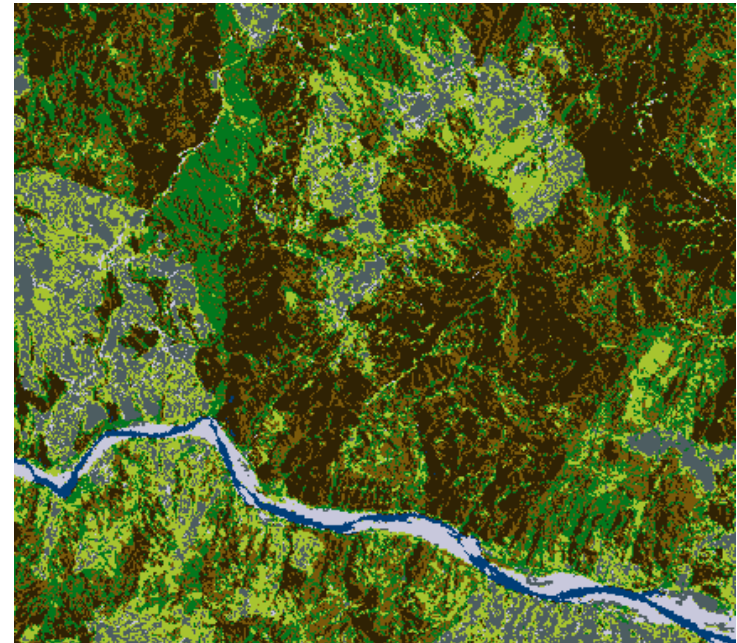
■ Ergebnis Originaldaten

- OAA: 42.6%
- Kappa: 0.31
- min(Cp.): 0.03%
- min(Cr.): 3.19%
- Dauer: 3.9 min



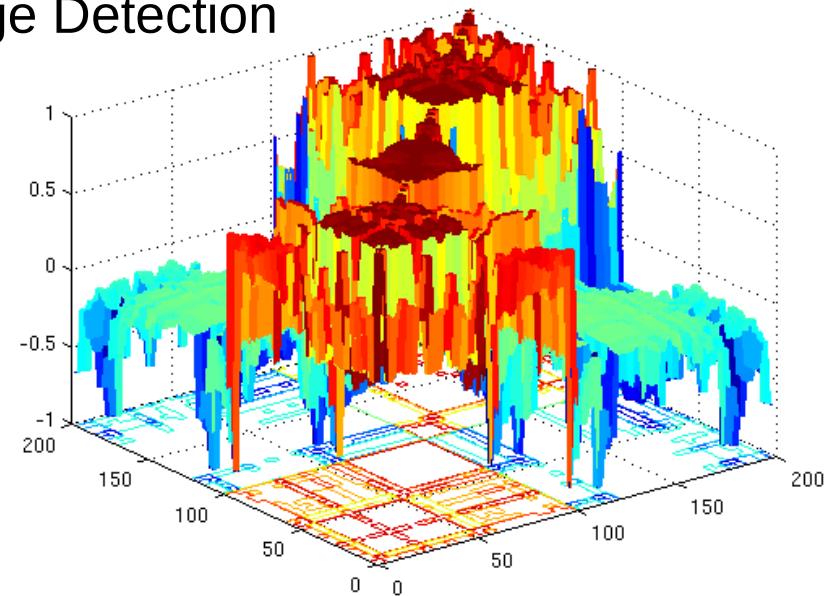
■ Ergebnis Originaldaten + Kbag

- OAA: 50.8%
- Kappa: 0.39
- min(Cp.): 34.8%
- min(Cr.): 34.6%
- Dauer: 2.7 min



Anwendungsbeispiel 3

Kernel Composition zur Change Detection



Kernel Composition Change Detection

- Camps-Valls et al. (2008) stellen ein ganzes kernel-basiertes Framework für multitemporale Ansätze und Change Detection Ansätze vor.
- Neben den „klassischen“ Composition Ansätzen (direct summation, weighted summation, cross-information kernel) werden der **difference kernel** und der **ratio kernel** vorgeschlagen
- Bsp: Difference Kernel:

$$K = K_{t_1}(x_i(t_1), x_j(t_1)) + K_{t_0}(x_i(t_0), x_j(t_0)) - K_{t_1, t_0}(x_i(t_1), x_j(t_0)) - K_{t_0, t_1}(x_i(t_0), x_j(t_1))$$
- Ähnlich wie difference image, aber im hochdimensionalen SVM Raum (RKHS)

Camps-Valls, G. & Gomez-Chova, L. & Muñoz-Mari, J. & Rojo-Alvarez J.L., TGRS 46(6), 2008, pp. 1822-1835

Beispiel 3: Change Detection Anwendung

- Ziel: Ausbreitung eines Steinbruchs zw. 2001 und 2005 beobachten
Braun et al. unpublished
- Verschiedene Zustände klassifizieren
 - Steinbruch, schon vorhanden in 2001
 - Steinbruch, zw. 2001 und 2005 neu
- Zwei Datensätze nutzen
 - Landsat ETM+ 2001
 - Landsat ETM+ 2005
- Dann SVM Klassifikation



2001



2005

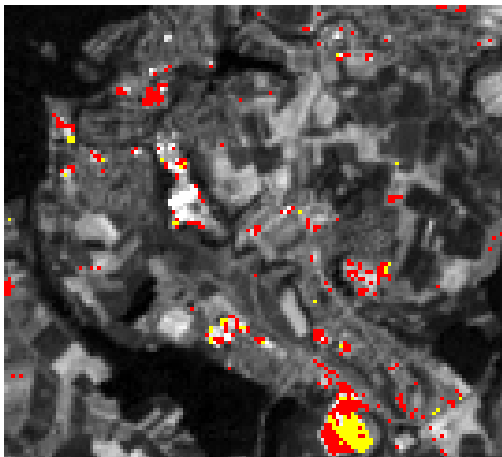


Groundtruth

Beispiel 3: Change Detection Anwendung

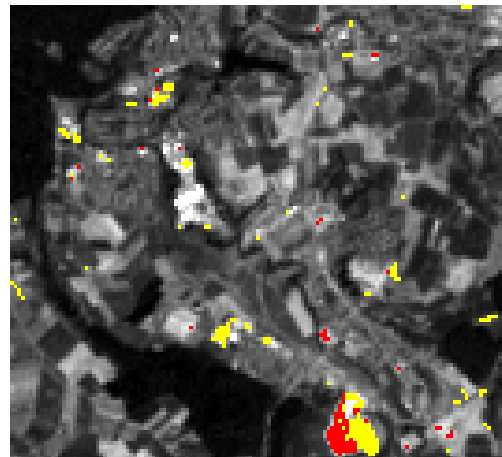
■ Post-Classification

■ OAA: 86,7%



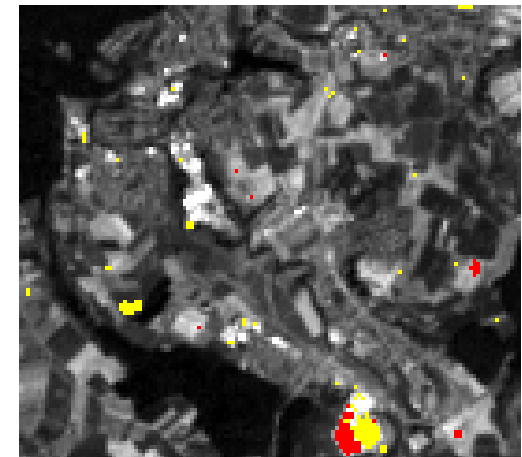
■ Pre-Classification (stacked)

■ OAA: 87,5%



■ Pre-Classification (Kernel-Comp.)

■ OAA: 88,6%



Überblick über die Literatur

Autoren	Jahr	Journal/Konf.	Vol(No)	Datenquelle 1	Datenquelle 2	Anwendungsschema
Tuia et al.	2011	JSTARS	4(1)	VHR	Cluster Kernel Kbag	Semi-Superv. Learning
Tan, Du	2011	COL ¹	9(1)	Hyperspectral	Wavelet	Spectral-Spatial
Chen et al.	2011	TGRS	49(11)	Hyperspectral	Spatial Correlation	Spectral-Spatial
Braun et al.	2011	ISPRS	--	Hyperspectral	LIDAR	Datafusion
Tuia et al.	2010	TGRS	48(10)	ein Kernel pro Bildkanal		Multi-Kernel Feat.Extr.
Tan, Du	2010	WHISPERS	--	Hyperspectral	Morph. Profile	Spectral-Spatial
Tuia et al.	2010	GRSL	7(1)	VHR	Morph. Profile	Spectral-Spatial
Huang et al.	2010	ICIEA ²	--	Hyperspectral	NWFE	Spectral-Spatial
Gomez-Chova et al.	2010	TGRS	48(1)	MERIS	Mean-Map K	Spectral-Spatial
Demir, Ertürk	2010	TGRS	48(11)	Hyperspectral	Empirical Mode Decomp	Spectral-Spatial
Tuia, Camps-Valls	2009	JURSE	--	VHR	Kbag	Spectral-Spatial SSL
Velasco-Forero, Manian	2009	GRSL	6(2)	Hyperspectral	Wavelet, Partial Diff.Eq	Spectral-Spatial
Demir, Ertürk	2009	IGARSS	--	Hyperspectral	Empirical Mode Decomp	Spectral-Spatial
Marconcini et al.	2009	GRSL	6(2)	Hyperspectral	Median Filter	Spectral-Spatial SSL
Valesco-Forero, Manian	2009	GRSL	6(2)	Hyperspectral	Wavelet, Partial Diff.Eq	Spectral-Spatial
Tuia, Camps-Valls	2009	GRSL	6(2)	Hyperspectral	Cluster Kernel Kbag	Spectral-Spatial SSL
Camps-Valls et al.	2008	TGRS	48(6)	Zeitpunkt 1	Zeitpunkt 2	Multitemporal, -source ³
Fauvel et al.	2008	ICIP ⁴	--	Hyperspectral	Self-Compl.Filt.	Spectral-Spatial
Copabianco, Camps-Valls	2008	SPIE	--	Hyperspectral	Correlation Context	Spectral-Spatial
Camps-Valls et al.	2007	TGRS	45(19)	Hyperspectral	Kontext (graph)	Spectral-Spatial SSL
Fauvel et al.	2007	IGARSS	--	VHR	Area Filter	Spectral-Spatial
Camps-Valls et al.	2006	SPIE	--	SAR	Multispectral	Datenfusion ³
Bandos et al.	2006	IGARSS	--	Hyperspectral	Kontext (graph)	Semi-Superv. Learning
Camps-Valls et al.	2005	GRSL	3(1)	Hyperspectral	Median Filter	Spectral-Spatial

¹Chinese Optics Letters

²IEEE Conference on Industrial Electronics and Applications

³weitere Ansätze im Paper vorhanden

⁴IEEE International Conference on Image Processing

Zusammenfassung

■ Vorteile

- Sehr flexible Methode der Fusion von Information
- Gewichtung möglich
- Zugeschnitten auf SVM
- Kein Hughes-Phänomen
- Oft höhere Genauigkeiten als andere Fusions-Ansätze

■ Nachteile

- v.a. Matlab: Man muss die Kernel Matrizen explizit im Workspace berechnen
- Memory Konflikte auch auf High-Performance Computern

- Die Methode ist noch nicht sehr etabliert, sehr viele Beiträge stammen direkt aus der Arbeitsgruppe Camps-Valls.
- Nicht limitiert auf SVM Klassifikatoren. Auch möglich mit IVM, RVM, etc...
- Für Anwender der SVM ein interessanter Ansatz

Vielen Dank für Ihre Aufmerksamkeit!

Andreas.Ch.Braun@Kit.edu