The Sparse Kernel Machine Model for Time Series Prediction

Laurentiu Bucur

Coordonator: Prof. Dr. Ing. Adina Magda Florea

Sparse Kernel Machines and High Performance Computing

Laurentiu Bucur

Coordonator: Prof. Dr. Ing. Adina Magda Florea

- Contributii la modelul de Masina Kernel Rara
- Aplicatii in finante si trafic

Cuprins

- Preliminarii: Predictie, Spatii Hilbert RKHS
- Kernel Ridge Regression
- Modele de Masini Kernel Rare (SKM):
  - SKM utilizand atractori
  - Masini suport vectoriale pentru regresie
- Contributie: Tehnici de predictie in haos. Studiu comparativ pentru serii de timp financiare.
  - Accent pe Masinile Kernel Rare
- Concluzii

Preliminarii

- Predictia seriilor de timp

Fie $\{y_n\}_{n \in Z}$ o serie de timp generata de:

$$y_n = f(x_1, x_2, \dots x_N)_n + \varepsilon_n$$

- f = componenta determinista necunoscuta

- $\{\varepsilon_n\}_{n \in Z}$ componenta stohastica : zgomot cu medie 0 si distributie de probabilitate $P_\varepsilon(\varepsilon)$. De obicei $P_\varepsilon(\varepsilon)$: distributia Gaussiana $G(0,\sigma^2)$ cu medie 0 si varianta σ^2.

Preliminarii

f se aproximeaza prin invatare cu $\hat{f}$:$R^K \rightarrow R$

aleasa dintr-o familie **F** de functii :

$$\hat{f} = \text{argmin}_{g \in F} \sum_{i=1}^{n} L(g(X_i), y_i)$$

- Doar K din N variabile ale modelului lui f pot fi **observabile**

Preliminarii

L(x,y) este o functie eroare:

- $L(x,y) = (x-y)^2$ eroarea patratica,

$$RSS = \sum_{i=n}^{m} (\hat{f}(X_i) - y_i)^2$$

$$sau$$

- $L(x,y) = \begin{cases} |x - y| - \varepsilon, if \ |x - y| \geq \varepsilon \\ 0, altfel \end{cases}$

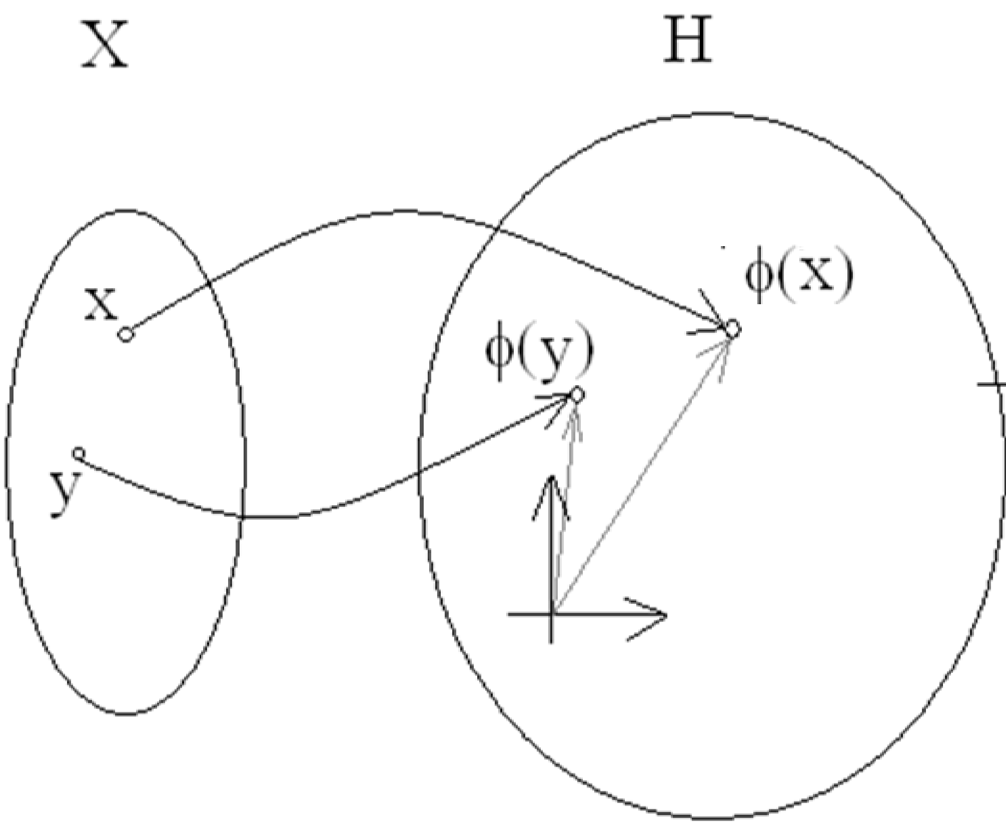
  ε-insensitive loss

Spatii Hilbert RKHS

- f se poate aproxima direct in spatiul observabil (de exemplu cu retele neuronale) sau in imaginea acestuia intr-un spatiu Hilbert de tip RKHS (cu masini kernel).

- Metode kernel: metode de ultima generatie, ulterioare anului 2000.
- Depasesc in general ca viteza si performanta alte modele de invatare statistica

Functii Kernel



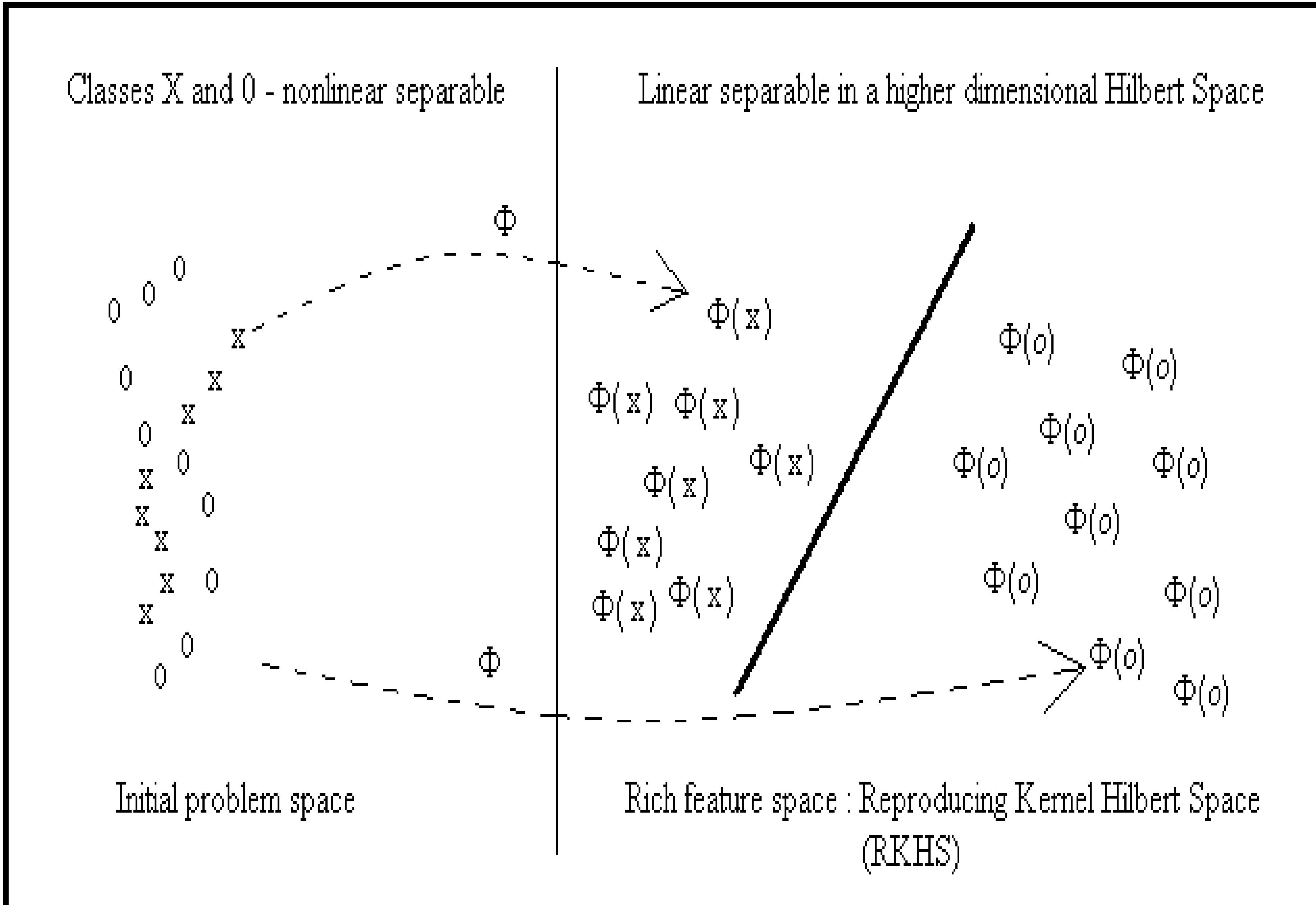
$K(x,y) = \langle \phi(x), \phi(y) \rangle$

K : X x X -> R
K pozitiv semi-definita
K- functie kernel
H - RKHS indus de K

Spatii Hilbert RKHS

- ‘Kernel trick’: nu trebuie evaluat produsul scalar in RKHS, ci evaluata doar functia kernel.

- Functiile kernel: utile in clasificare si predictie:
  - Clase neseparabile in spatiul initial X sunt separabile in RKHS
  - O suprafata $g(x,y,z\dots) = 0$ in X se mapeaza intr-un hiperplan $h(\phi(x), \phi(y),\dots) = 0$ in RKHS, deoarece spatiile Hilbert sunt infinit dimensionale


Classes X and 0 - nonlinear separable
Linear separable in a higher dimensional Hilbert Space
Φ
Φ(x)
Φ(o)
Initial problem space
Rich feature space : Reproducing Kernel Hilbert Space (RKHS)

RKHS al unei functii kernel

- Fie X o multime (posibil) infinit numarabila
- K : X x X → R, functie simetrica pozitiv semi-definita
- Teorema **Moore-Aronsajn** : Pentru K exista un RKHS pentru care K este Reproducing Kernel
- RKHS indus de K peste X este:

$$\mathcal{H} = \{\sum_{i=1}^{l} \alpha_i K(x_i, \cdot) \ : l \in \boldsymbol{N}, x_i \in X, \alpha_i \in \boldsymbol{R}, i = 1 \dots l\}$$

cu produsul scalar

$$\langle f, g \rangle = \sum_{i=1}^{l} \sum_{j=1}^{n} \alpha_i \beta_j K(x_i, z_j)$$

Kernel Ridge Regression

- Cel mai simplu model din familia de metode kernel pentru predictie
- Se folosesc toate perechile etichetate (x,y) dintr-o multime finita si numarabila X x R, Card (X)=N, pentru invatare supervizata
- Din familia de functii:

$$\mathcal{H}_{\mathbf{N}} = \left\{ \sum_{i=1}^{N} \alpha_i K(x_i, \cdot) \quad , x_i \in X, \alpha_i \in \boldsymbol{R}, i = 1 \ldots N \right\}$$

se calculeaza vectorul f dat de $\alpha = \{\alpha_i\}_{i=1..N}$

Kernel Ridge Regression

$$\alpha = (G + \lambda I_N)^{-1} y$$

G este matricea Gram (Kernel) N x N;

$$G_{ij} =< \Phi(x_i), \Phi(x_j) > = K(x_i, x_j) \quad \text{i,j=1..N}$$

$\lambda>0$ – constanta de regularizare, aleasa apriori

Pentru calculul vectorului α matricea se inverseaza utilizand metoda Conjugate Gradient, deoarece matricea este pozitiv definita si inversabila.

Ridge Regression : Probleme

- Timpul de antrenare este O(N^3)
- Pentru N suficient de mare, evaluarea masinii kernel in timp real este prohibitiva

- Kernel Ridge Regression – model generativ: se folosesc toate datele disponibile pentru a modela o distributie de probabilitate
- Clasa de metode: Kernel Density Estimation

Masini Kernel Rare

- Este nevoie de masini kernel cu mai putini vectori baza decat modelele generative.
- Masini Kernel Rare (Sparse Kernel Machines):

$$\mathrm{g(x)} = \sum_{i=1}^{m} \beta_i K(z_i, x) + c$$

- Exista in subspatii m - dimensionale ale spatiului Hilbert $\mathcal{H}$ indus de K
- m << N – eficienta in stocare si executie
- Complexitate Rademacher (capacitate) redusa

Masini Kernel Rare

$$g(x) = \sum_{i=1}^{m} \beta_i K(z_i, x) + c$$

- $c, \beta_i \in R$, $i=1..m$ – ponderi

- $\{z_i\}_{i=1,m}$ - Setul Redus (Reduced Set - RS) al lui g

  $z_i \in X$, $i=1..m$

Masini Kernel Rare

1) Se calculeaza modelul generativ *f* utilizand Kernel Ridge Regression

2) Se determina *g* (RS si ponderile) astfel incat sa se minimizeze distanta:

$$\|\text{f-g}\|^2 = \sum_{i,j=1}^{N} \alpha_i \alpha_j K(x_i, x_j) + \sum_{i,j=1}^{m} \beta_i \beta_j K\left(z_i, z_j\right) - 2 \sum_{i=1}^{N} \sum_{j=1}^{m} \alpha_i \beta_j K(x_i, z_j)$$

– Masina Kernel Rara – proiectia masinii kernel *f* pe subspatiul *Span(RS)*

Calculul Masinilor Kernel Rare

- Calculul RS utilizand atractori in RKHS (Lee, Jung, Lee, 2009)
- Masini Suport Vectoriale pentru Regresie (Bishop, 2006)
- Metode noi pentru calculul RS, **principalele contributii ale tezei**

Calculul RS utilizand atractori

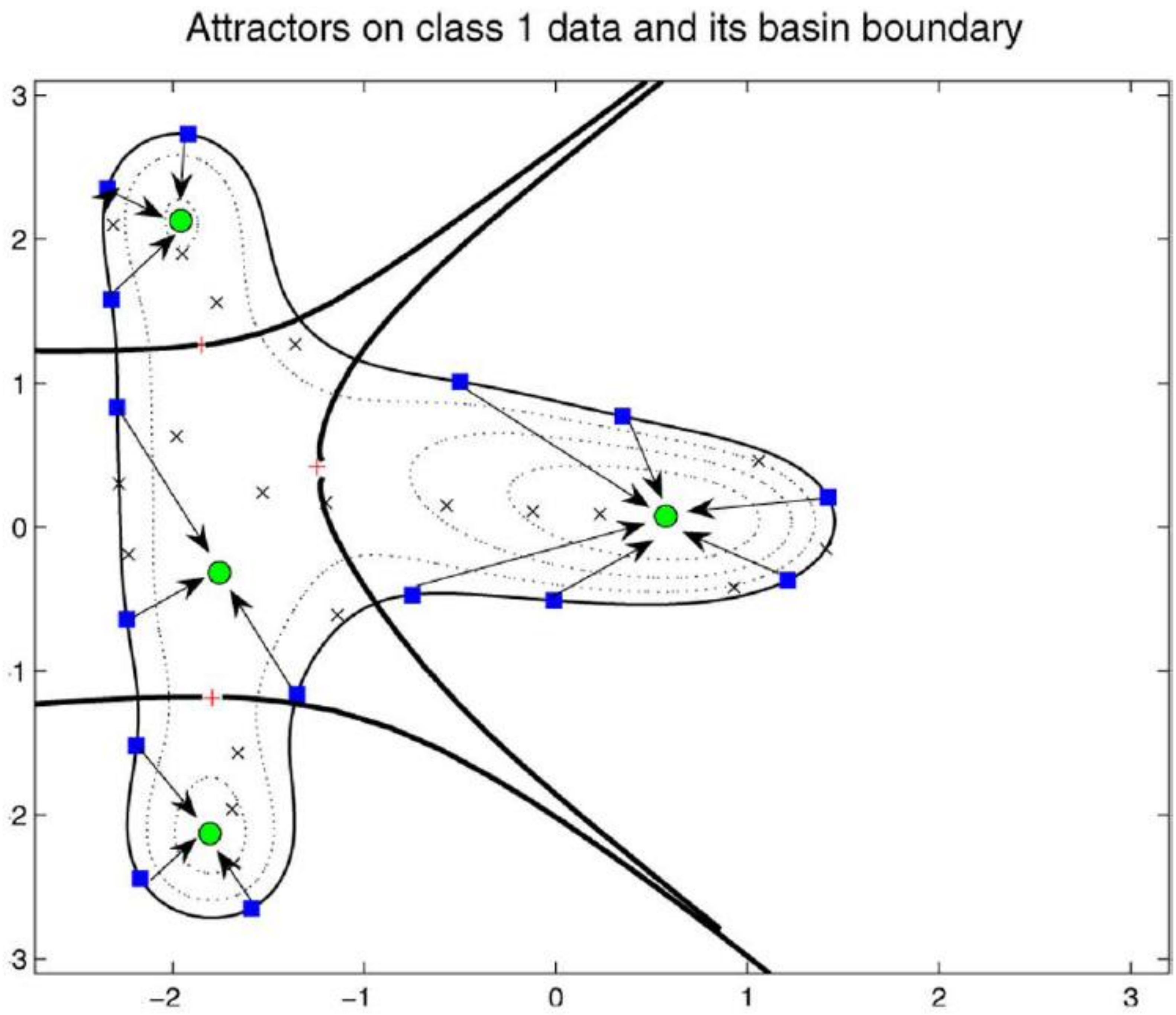
- Se calculeaza *f* utilizand Kernel Ridge Regression
- Pentru fiecare kernel K(x,.) din expresia lui f:

  Se calculeaza prin integrare numerica punctul stationar z(x) al ecuatiei diferentiale:

$$\frac{dx}{dt} = -\nabla\rho(x)$$

unde:

$$\rho(x) = \|f - \Phi(x)\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j K\left(x_i, x_j\right) + K(x,x) - 2\sum_{j=1}^{N} \alpha_i K(x_i, x)$$



Attractors on class 1 data and its basin boundary
3
2
1
0
-1
-2
-3
−2
−1
0
1
2
3

SKM utilizand atractori

- Fiecare punct din X apartine unui singur bazin de atractie
- Centrii distincti ai bazinelor de atractie obtinuti prin integrare numerica corespund elementelor z_i , $i=1..m$ si alcatuiesc Setul Redus (RS) al masinii kernel rare.
- Lee, Jung, Lee nu ofera o metoda obiectiva de identificare a centrilor bazinelor de atractie
- Am ales un subset ortogonal din multimea $\{z(x_i)\}_{i=1..N}$ obtinuta prin integrare numerica:

- $K(z_i, z_j) = 0$, $i,j=1..m$
- numarul m de centri din RS depinde de:
  - Functia kernel K
  - Numarul si pozitia punctelor $\{x_i\}_{i=1..N}$ din masina kernel *f*

Problema: Nr. de pasi din integrarea numerica este necunoscut

- De obicei K este functia Kernel Gaussian

Calculul ponderilor SKM

$$\beta = (K^{zz})^{-1} K^{zx} \alpha$$

unde:

- $\alpha = (\alpha_1, \alpha_2, \alpha_3, ..\alpha_N)^T$ – vectorul pondere al masinii f
- K^{zz} - matrice kernel $M \; x \; M$, $(K^{zz})_{ij} = K(z_i, z_j)$
- K^{zx} - matrice kernel $M \; x \; N$, $(K^{zx})_{ij} = K(z_i, x_j)$
- $\{x_i\}_{i=1..N} \subseteq X$ punctele din multimea de antrenament utilizate in expresia lui f

Masini Suport Vectoriale pentru Regresie

- **Reduced Set**: multimea punctelor $\{x_i\}$ din setul de antrenament pentru care $|f(x_i)-y_i| \geq \varepsilon$ sunt vectori suport pentru regresie. ε ales apriori
- Cu cat ε este mai mare cu atat nr. vectori suport este mai mic
- **Ponderile** se calculeaza utilizand optimizare convexa cu constrangeri (conditiile Karush-Kuhn-Tucker)
- Forma duala a ponderilor
- Implementare in pachetul LIBSVM

Contributia #1 (UPB Sci. Bulletin, 2011)

- **Studiu comparativ** asupra performantei de predictie a metodelor neliniare existente, inclusiv masinile kernel rare, pe serii de timp financiare
- **Notorii pentru dificultate** : haos determinist + zgomot
- **8 perechi valutare**
- **2 indicatori ai volatilitatii** din Chicago Board of Options Exchange (CBOE): Volatility Index (**fear gauge**) si **Put-to-call Ratio**

$\approx$ 81000 inregistrari

Metode analizate

- Nearest Neighbours (Hannias, Karras, 2007)
- Nearest Trajectory (McNames, 1998, 2002)
- Multiple Backprop Multi-Layer Perceptron
  - Spatiul fazelor brut
  - Spatiul fazelor seriei de timp diferenta
- Kernel Ridge Regression
- Masini Suport Vectoriale pentru regresie (SKM)
- Masini Kernel Rare utilizand atractori (Lee, Jung, Lee, 2009)

Data set no.	Original series (symbol and time frame)	Best RMSE algorithm	**Best RMSE**	Best directional predictor	Best directional accuracy (%)
1	VIX Daily	Kernel-Ridge	**1.247**	MBP-Global	**56.12***
2		Kernel-Ridge, SKM-Attractors	**1.625**	ε-SVR	**55.04***
3	Put-To-Call ratio Daily	ε-SVR	**0.273**	ε-SVR	**67.26***
4		ε-SVR	**0.268**	MBP-Global	**68.81***
5	EUR/USD 4h	Kernel-Ridge	**0.003**	MBP-Local	50.94
6		Kernel-Ridge	**0.003**	Kernel-Ridge	50.56
7	GBP/USD Daily	SKM-Attractors	**0.0097**	SKM-Attractors	50.86
8		Kernel-Ridge	**0.009**	MBP-Global	**51.51****
9	GBP/USD 4h	MBP-Global	**0.0049**	MBP-Global	50.78
10		MBP-Global	**0.0049**	SKM-Attractors	**51.09****
11	USD/CAD Daily	Nearest-Neighbours	**0.0062**	Nearest-Neighbours	**51.548****
12		SKM-Attractors, MBP-Global	**0.0073**	MBP-Global	**52.75****

Concluzii

- **Masinile Kernel Rare:**
  - Metode parametrice de predictie
  - Viteza sporita de executie
  - Comparabile ca performanta cu metode existente (retele neuronale, metode kernel dedicate problemei predictiei)

Contributie: Studiu exhaustiv al unora dintre cei mai buni algoritmi de predictie neliniara, in paralel cu modelele recent introduse (2006-2009)

Concluzii

- Grad inalt de predictibilitate a seriilor de timp Volatility Index (Fear gauge) si Put-to-call ratio
- Grad scazut de predictibilitate a seriilor valutare – pretul de inchidere la 4 ore si zilnic
- Rezultatele analizei state of the art (metode kernel, retele neuronale): performanta **<u>nesatisfacatoare</u>** pentru seriile de timp valutare
- Haos + zgomot : provocari noi, nevoie de abordari si rezultate noi in metodele kernel

Concluzii

- Cele mai mari sanse de reusita: imbunatatirea modelelor de masini kernel rare
- Necesara utilizarea High Performance Computing pentru rezultate optime
- Abordare interdisciplinara:
  - Procesarea Semnalelor
  - Machine Learning
  - Statistica