Data visualization and dimensionality reduction using kernel maps with a reference point

Johan Suykens

K.U. Leuven, ESAT-SCD/SISTA
Kasteelpark Arenberg 10
B-3001 Leuven (Heverlee), Belgium

Tel: 32/16/32 18 02 - Fax: 32/16/32 19 70
Email: johan.suykens@esat.kuleuven.be
http://www.esat.kuleuven.be/scd/

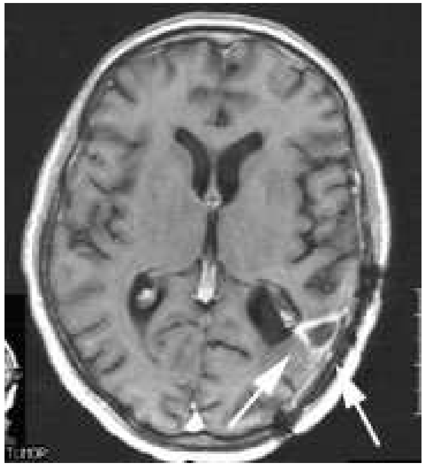
International Conference on Computational Harmonic Analysis
Shanghai June 2007

Contents

- Context: support vector machines and kernel based learning
- Core problems: least squares support vector machines
- Classification and kernel principal component analysis
- Data visualization
- Kernel eigenmap methods
- Kernel maps with a reference point: linear system solution
- Examples

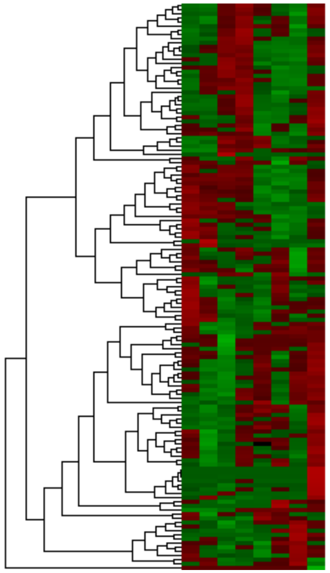
Living in a data world

biomedical

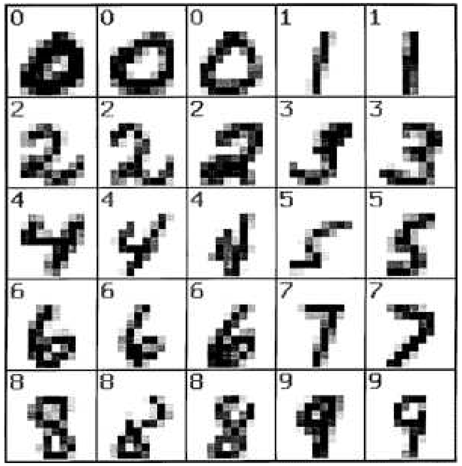


energy

process industry



bio-informatics



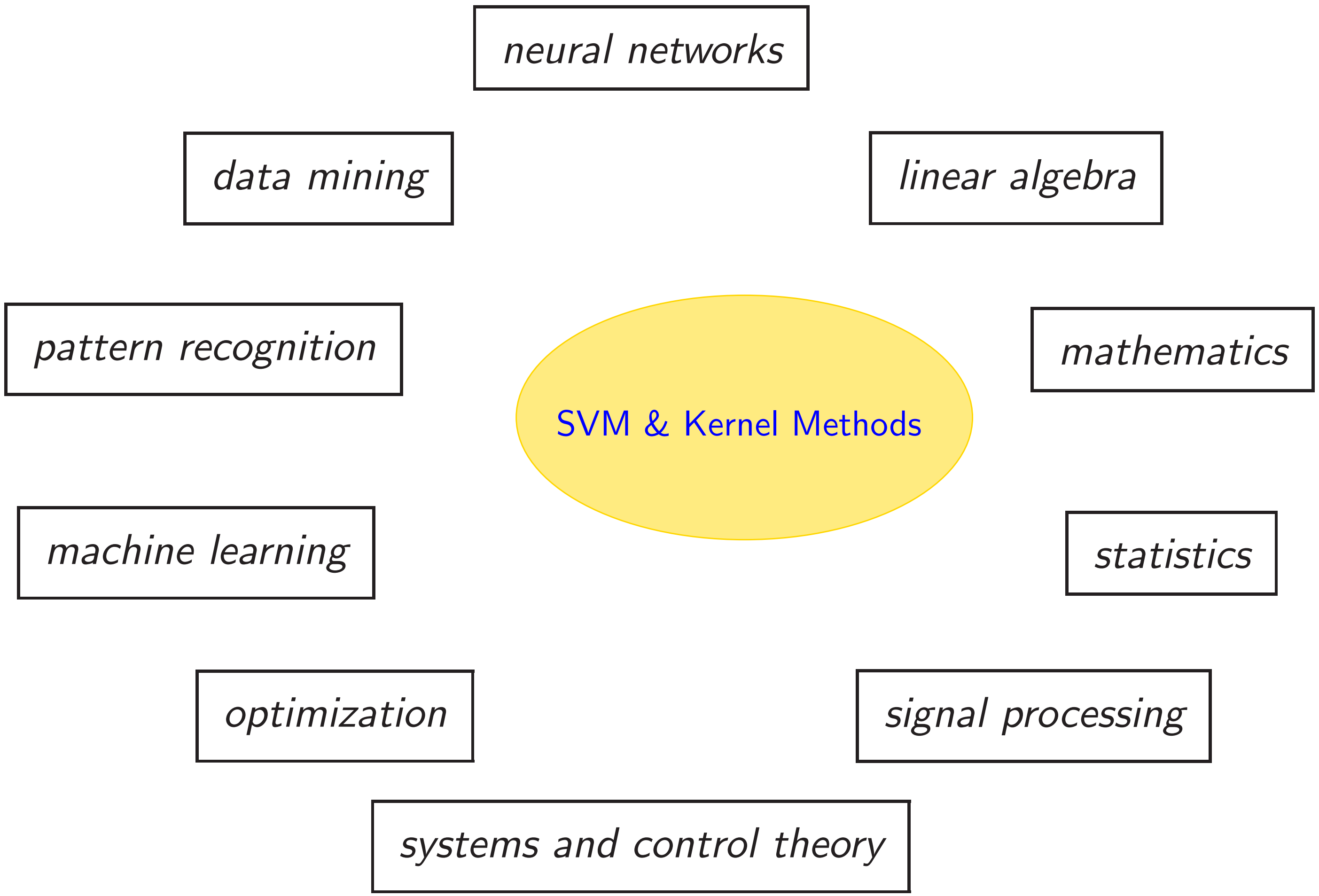
multimedia

traffic

Support vector machines and kernel methods: context

- With new technologies (e.g. in microarrays, proteomics) massive data sets become available that are high dimensional.
- Tasks and objectives: predictive modelling, knowledge discovery and integration, data fusion (classification, feature selection, prior knowledge incorporation, correlation analysis, ranking, robustness).
- Supervised, unsupervised or semi-supervised learning depending on the given data and problem.
- Need for modelling techniques that are able to operate on different data types (sequences, graphs, numerical, categorical, ...)
- Linear as well as nonlinear models
- Reliable methods: numerically, computationally, statistically

Kernel based learning: interdisciplinary challenges



Estimation in Reproducing Kernel Hilbert Spaces (RKHS)

- Variational problem: [Wahba, 1990; Poggio & Girosi, 1990; Evgeniou et al., 2000] find function f such that

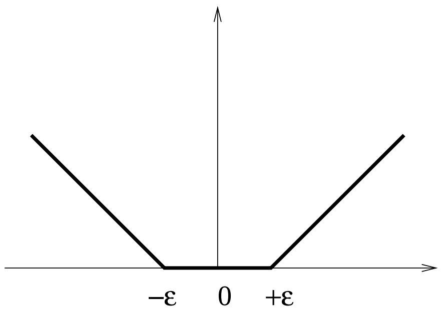
$$\min_{f \in \mathcal{H}} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i)) + \lambda \|f\|_K^2$$

with $L(\cdot, \cdot)$ the loss function. $\|f\|_K$ is norm in RKHS $\mathcal{H}$ defined by K.

- Representer theorem: for convex loss function, solution of the form

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i)$$

Reproducing property $f(x) = \langle f, K_x \rangle_K$ with $K_x(\cdot) = K(x, \cdot)$



- Some special cases:
$L(y, f(x)) = (y - f(x))^2$: regularization network
$L(y, f(x)) = |y - f(x)|_\epsilon$: SVM regression with ϵ-insensitive loss function

Different views on kernel based models

SVM

LS–SVM

Kriging

RKHS

Gaussian Processes

Some early history on RKHS:

1910-1920: Moore

1940: Aronszajn

1951: Krige

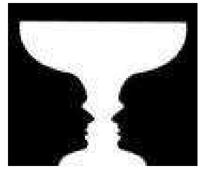
1970: Parzen

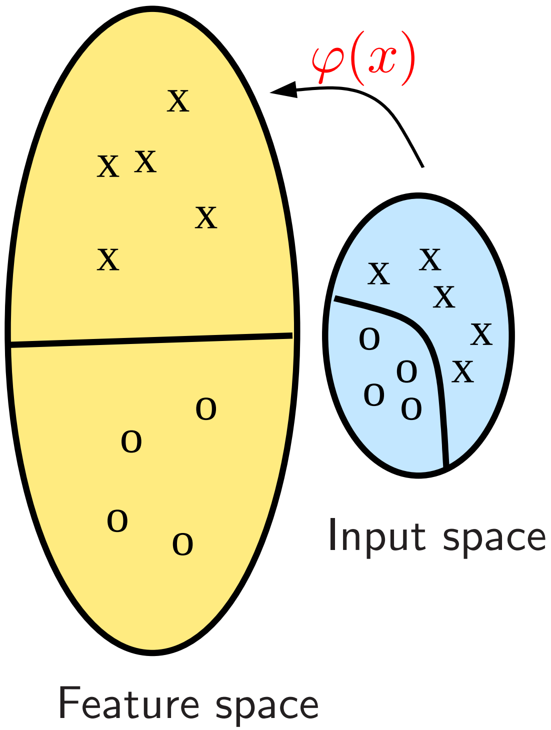
1971: Kimeldorf & Wahba

Obtaining complementary insights from different perspectives:
kernels are used in different methodologies

Support vector machines (SVM):	optimization approach (primal/dual)
Reproducing kernel Hilbert spaces (RKHS):	variational problem, functional analysis
Gaussian processes (GP):	probabilistic/Bayesian approach

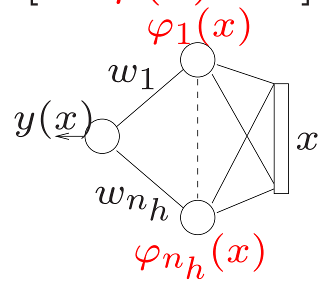
SVMs: living in two worlds ...





Primal space:

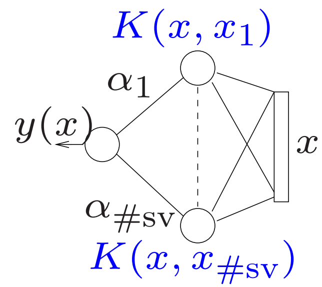
$$y(x) = \text{sign}[w^T \varphi(x) + b]$$



$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$$ **("Kernel trick")**

Dual space:

$$y(x) = \text{sign}[\sum_{i=1}^{\#\text{sv}} \alpha_i y_i K(x, x_i) + b]$$



Least Squares Support Vector Machines: “core problems”

- Regression (RR)

$$\min_{w,b,e} w^T w + \gamma \sum_i e_i^2 \quad \text{s.t.} \quad y_i = w^T \varphi(x_i) + b + e_i, \quad \forall i$$

- Classification (FDA)

$$\min_{w,b,e} w^T w + \gamma \sum_i e_i^2 \quad \text{s.t.} \quad y_i (w^T \varphi(x_i) + b) = 1 - e_i, \quad \forall i$$

- Principal component analysis (PCA)

$$\min_{w,b,e} -w^T w + \gamma \sum_i e_i^2 \quad \text{s.t.} \quad e_i = w^T \varphi(x_i) + b, \quad \forall i$$

- Canonical correlation analysis/partial least squares (CCA/PLS)

$$\min_{w,v,b,d,e,r} w^T w + v^T v + \nu_1 \sum_i e_i^2 + \nu_2 \sum_i r_i^2 - \gamma \sum_i e_i r_i \text{ s.t. } \left\{ \begin{array}{lcl} e_i & = & w^T \varphi_1(x_i) + b \\ r_i & = & v^T \varphi_2(y_i) + d \end{array} \right.$$

- partially linear models, spectral clustering, subspace algorithms, ...

LS-SVM classifier

- Preserve support vector machine [Vapnik, 1995] methodology, but simplify via least squares and equality constraints [Suykens, 1999]

- **Primal problem:**

$$\min_{w,b,e} \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{i=1}^{N} e_i^2 \;\; \text{such that} \;\; y_i \, [w^T \varphi(x_i) + b] = 1 - e_i, \;\; i = 1, ..., N$$

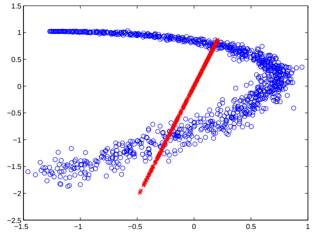
- **Dual problem:**

$$\left[\begin{array}{c|c} 0 & y^T \\ \hline y & \Omega + I/\gamma \end{array} \right] \left[\begin{array}{c} b \\ \hline \alpha \end{array} \right] = \left[\begin{array}{c} 0 \\ \hline 1_N \end{array} \right]$$

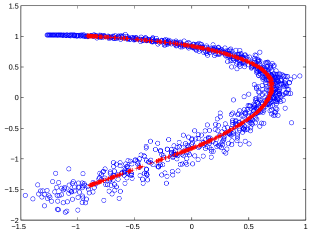
where $\Omega_{ij} = y_i y_j \, \varphi(x_i)^T \varphi(x_j) = y_i y_j \, K(x_i, x_j)$ and $y = [y_1; ...; y_N]$.

- LS-SVM classifiers perform very well on 20 UCI data sets [Van Gestel et al., ML 2004]
Winning results in competition WCCI 2006 by [Cawley, 2006]

Kernel PCA: primal and dual problem



linear PCA



kernel PCA (RBF kernel)

- Primal problem: [Suykens et al., 2003]

$$\min_{w,b,e} \; -\frac{1}{2}w^T w + \frac{1}{2}\gamma \sum_{i=1}^{N} e_i^2 \;\; \text{such that} \;\; e_i = w^T \varphi(x_i) + b, \; i = 1, ..., N.$$

- KPCA [Schölkopf et al., 1998]: Dual problem = kernel PCA:

$$\Omega_c \alpha = \lambda \alpha \;\; \text{with} \;\; \lambda = 1/\gamma$$

with $\Omega_{c,ij} = (\varphi(x_i) - \hat{\mu}_\varphi)^T (\varphi(x_j) - \hat{\mu}_\varphi)$ the *centered kernel matrix*.
Underlying LS-SVM model allows to make out-of-sample extensions.

Core models + additional constraints

- Monoticity constraints: [Pelckmans et al., 2005]

$$\min_{w,b,e} \; w^T w + \gamma \sum_{i=1}^{N} e_i^2 \;\; \text{s.t.} \;\; \begin{cases} y_i = w^T \varphi(x_i) + b + e_i, & (i = 1, \ldots, N) \\ w^T \varphi(x_i) \leq w^T \varphi(x_{i+1}), & (i = 1, \ldots, N-1) \end{cases}$$

- Structure detection: [Pelckmans et al., 2005; Tibshirani, 1996]

$$\min_{w,e,t} \; \rho \sum_{p=1}^{P} t_p + \sum_{p=1}^{P} {w^{(p)}}^T w^{(p)} + \gamma \sum_{i=1}^{N} e_i^2 \;\; \text{s.t.} \;\; \begin{cases} y_i = \sum_{p=1}^{P} {w^{(p)}}^T \varphi^{(p)}(x_i^{(p)}) + e_i, & (\forall i) \\ -t_p \leq {w^{(p)}}^T \varphi^{(p)}(x_i^{(p)}) \leq t_p, & (\forall i, \forall p) \end{cases}$$

- Autocorrelated errors: [Espinoza et al., 2006]

$$\min_{w,b,r,e} \; w^T w + \gamma \sum_{i=1}^{N} r_i^2 \;\; \text{s.t.} \;\; \begin{cases} y_i = w^T \varphi(x_i) + b + e_i, & (i = 1, .., N) \\ e_i = \rho e_{i-1} + r_i, & (i = 2, \ldots, N) \end{cases}$$

- Spectral clustering: [Alzate & Suykens, 2006; Chung, 1997; Shi & Malik, 2000]

$$\min_{w,b,e} \; -w^T w + \gamma e^T D^{-1} e \;\; \text{s.t.} \;\; e_i = w^T \varphi(x_i) + b, \;\; (i = 1, \ldots, N)$$

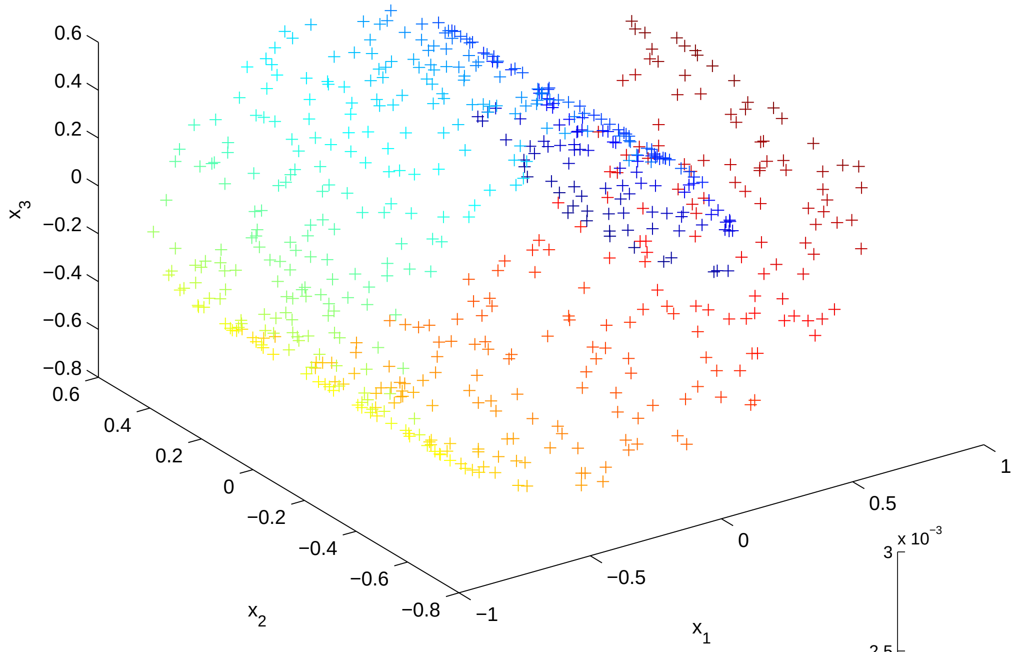
Dimensionality reduction and data visualization

- Traditionally:
  commonly used techniques are e.g. principal component analysis, multi-dimensional scaling, self-organizing maps

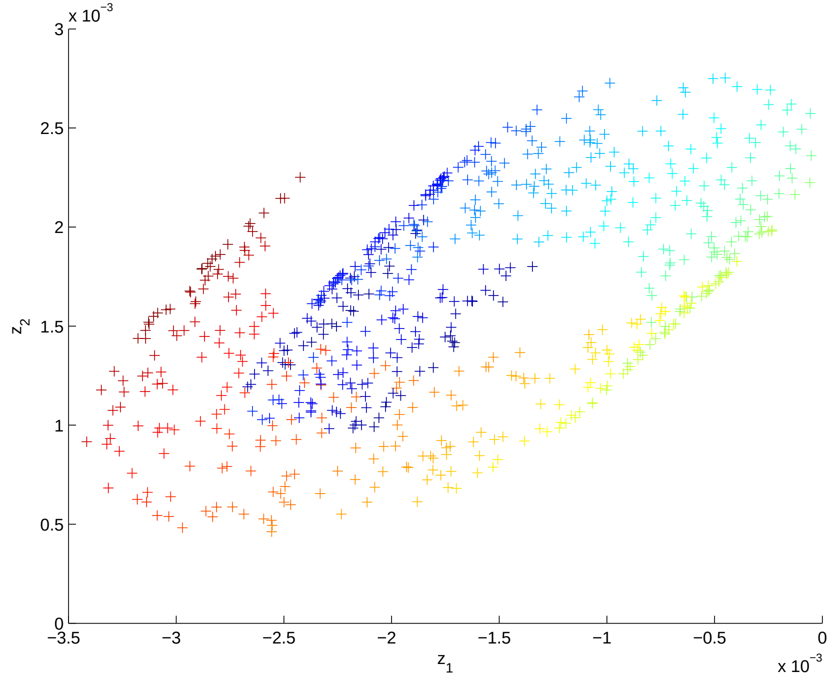
- More recently:
  isomap, locally linear embedding, Hessian locally linear embedding, diffusion maps, Laplacian eigenmaps
  ("kernel eigenmap methods and manifold learning")
  [Roweis & Saul, 2000; Coifman et al., 2005; Belkin et al., 2006]

- Relevant issues:
  - *learning and generalization* [Cucker & Smale, 2002; Poggio et al., 2004]
  - *model representations and out-of-sample extensions*
  - *convex/non-convex problems, computational complexity* [Smale, 1997]

- **Kernel maps with reference point** (KMref) [Suykens, 2007]:
  data visualization and dimensionality reduction by solving linear system



(3D given)



(2D KMref result)

A criterion related to locally linear embedding

- Given training data set $\{x_i\}_{i=1}^N$ with $x_i \in \mathbb{R}^p$.
  Dimensionality reduction to $\{z_i\}_{i=1}^N$ with $z_i \in \mathbb{R}^d$ ($d = 2$ or $d = 3$).

- Objective

$$\min_{z_i \in \mathbb{R}^d} \; -\frac{\gamma}{2} \sum_{i=1}^N \|z_i\|_2^2 + \frac{1}{2} \sum_{i=1}^N \|z_i - \sum_{j=1}^N s_{ij} z_j\|_2^2$$

  where e.g.

$$s_{ij} = \exp(-\|x_i - x_j\|_2^2 / \sigma^2)$$

- Solution follows from eigenvalue problem

$$Rz = \gamma z$$

  with $z = [z_1; z_2; ...; z_N]$ and $R = (I - P)^T (I - P)$ where $P = [s_{ij} I_d]$.

Introducing a core model

- Realize the nonlinear mapping $x \mapsto z$ through a least squares support vector machine regression:

$$\min_{z,w_j,e_{i,j}} \quad -\frac{\gamma}{2} z^T z + \frac{1}{2}(z - Pz)^T (z - Pz) + \frac{\nu}{2} \sum_{j=1}^{d} w_j^T w_j + \frac{\eta}{2} \sum_{i=1}^{N} \sum_{j=1}^{d} e_{i,j}^2$$
$$\text{such that} \quad c_{i,j}^T z = w_j^T \varphi_j(x_i) + e_{i,j}, \quad \forall i = 1, ..., N; j = 1, ..., d$$

- Primal model representation with evaluation at point $x^* \in \mathbb{R}^p$:

$$\hat{z}_{*,j} = w_j^T \varphi_j(x^*)$$

with $w_j \in \mathbb{R}^{n_{h_j}}$ and feature maps $\varphi_j(\cdot) : \mathbb{R}^p \to \mathbb{R}^{n_{h_j}}$ $(j = 1, ..., d)$

Kernel maps and eigenvalue problem

- Solution follows from **eigenvalue problem**, e.g. for $d = 2$:

$$\left(R + V_1(\frac{1}{\nu}\Omega_1 + \frac{1}{\eta}I)^{-1}V_1^T + V_2(\frac{1}{\nu}\Omega_2 + \frac{1}{\eta}I)^{-1}V_2^T \right) z = \gamma z$$

with kernel matrices Ω_1, Ω_2:

$$\begin{array}{l} \Omega_{1,ij} = K_1(x_i, x_j) = \varphi_1(x_i)^T \varphi_1(x_j) \\ \Omega_{2,ij} = K_2(x_i, x_j) = \varphi_2(x_i)^T \varphi_2(x_j) \end{array}$$

matrices V_1, V_2:

$$V_1 = [c_{1,1}\, c_{2,1} \ldots c_{N,1}], V_2 = [c_{1,2}\, c_{2,2} \ldots c_{N,2}]$$

- However, **selection** of the best solution from this pool of $2N$ candidates is **not straightforward** (the best solution is not necessarily given by the largest or smallest eigenvalue here).

Kernel maps with reference point: problem statement

- Kernel maps with reference point:
  - **LS-SVM core part**: realize dimensionality reduction $x \mapsto z$
  - **reference point** q (e.g. first point; sacrificed in the visualization)

- Example: $d = 2$

$$\min_{z,w_1,w_2,b_1,b_2,e_{i,1},e_{i,2}} \quad \frac{1}{2}(z - P_D z)^T (z - P_D z) + \frac{\nu}{2}(w_1^T w_1 + w_2^T w_2) + \frac{\eta}{2}\sum_{i=1}^{N}(e_{i,1}^2 + e_{i,2}^2)$$

$$\text{such that} \quad \begin{array}{l} c_{1,1}^T z = q_1 + e_{1,1} \\ c_{1,2}^T z = q_2 + e_{1,2} \\ c_{i,1}^T z = w_1^T \varphi_1(x_i) + b_1 + e_{i,1}, \quad \forall i = 2, ..., N \\ c_{i,2}^T z = w_2^T \varphi_2(x_i) + b_2 + e_{i,2}, \quad \forall i = 2, ..., N \end{array}$$

Coordinates in low dimensional space: $z = [z_1; z_2; ...; z_N] \in \mathbb{R}^{dN}$
Regularization term: $(z - P_D z)^T (z - P_D z) = \sum_{i=1}^{N} \| z_i - \sum_{j=1}^{N} s_{ij} D z_j \|_2^2$
with D diagonal matrix and $s_{ij} = \exp(-\|x_i - x_j\|_2^2 / \sigma^2)$.

Kernel maps with reference point: solution

- The **unique solution** to the problem is given by the **linear system**

$$\left[\begin{array}{c|c|c} U & -V_1 M_1^{-1} 1 & -V_2 M_2^{-1} 1 \\ \hline -1^T M_1^{-1} V_1^T & 1^T M_1^{-1} 1 & 0 \\ \hline -1^T M_2^{-1} V_2^T & 0 & 1^T M_2^{-1} 1 \end{array}\right] \left[\begin{array}{c} z \\ \hline b_1 \\ \hline b_2 \end{array}\right] = \left[\begin{array}{c} \eta(q_1 c_{1,1} + q_2 c_{1,2}) \\ \hline 0 \\ \hline 0 \end{array}\right]$$

with matrices

$$U = (I - P_D)^T (I - P_D) - \gamma I + V_1 M_1^{-1} V_1^T + V_2 M_2^{-1} V_2^T + \eta c_{1,1} c_{1,1}^T + \eta c_{1,2} c_{1,2}^T$$
$$M_1 = \frac{1}{\nu}\Omega_1 + \frac{1}{\eta} I \, , \; M_2 = \frac{1}{\nu}\Omega_2 + \frac{1}{\eta} I$$
$$V_1 = [c_{2,1} \ldots c_{N,1}] \, , \; V_2 = [c_{2,2} \ldots c_{N,2}]$$

kernel matrices $\Omega_1, \Omega_2 \in \mathbb{R}^{(N-1)\times(N-1)}$:
$\Omega_{1,ij} = K_1(x_i, x_j) = \varphi_1(x_i)^T \varphi_1(x_j)$, $\Omega_{2,ij} = K_2(x_i, x_j) = \varphi_2(x_i)^T \varphi_2(x_j)$
positive definite kernel functions $K_1(\cdot,\cdot), K_2(\cdot,\cdot)$.

Kernel maps with reference point: model representations

- The **primal and dual model representations** allow making **out-of-sample extensions**. Evaluation at point $x^* \in \mathbb{R}^p$:

$$\begin{aligned} \hat{z}_{*,1} &= w_1^T \varphi_1(x^*) + b_1 &= \frac{1}{\nu} \sum_{i=2}^{N} \alpha_{i,1} K_1(x_i, x^*) + b_1 \\ \hat{z}_{*,2} &= w_2^T \varphi_2(x^*) + b_2 &= \frac{1}{\nu} \sum_{i=2}^{N} \alpha_{i,2} K_2(x_i, x^*) + b_2 \end{aligned}$$

  Estimated coordinates for visualization: $\hat{z}_* = [\hat{z}_{*,1}; \hat{z}_{*,2}]$.

- $\alpha_1, \alpha_2 \in \mathbb{R}^{N-1}$ are the unique solutions to the **linear systems**

$$M_1 \alpha_1 = V_1^T z - b_1 1_{N-1} \text{ and } M_2 \alpha_2 = V_2^T z - b_2 1_{N-1}$$

  and $\alpha_1 = [\alpha_{2,1}; ...; \alpha_{N,1}]$, $\alpha_2 = [\alpha_{2,2}; ...; \alpha_{N,2}]$, $1_{N-1} = [1; 1; ...; 1]$.

Proof - Lagrangian

- Only equality constraints: **optimal model representation and solution is obtained in a systematic and straightforward way.**

- Lagrangian:

$$\mathcal{L}(z, w_1, w_2, b_1, b_2, e_{i,1}, e_{i,2};\ \beta_{1,1}, \beta_{1,2}, \alpha_{i,1}, \alpha_{i,2}) =$$
$$-\frac{\gamma}{2} z^T z + \frac{1}{2}(z - P_D z)^T (z - P_D z) + \frac{\nu}{2}(w_1^T w_1 + w_2^T w_2) +$$
$$\frac{\eta}{2} \sum_{i=1}^N (e_{i,1}^2 + e_{i,2}^2) + \beta_{1,1}(c_{1,1}^T z - q_1 - e_{1,1}) + \beta_{1,2}(c_{1,2}^T z - q_2 - e_{1,2})$$
$$+ \sum_{i=2}^N \alpha_{i,1}(c_{i,1}^T z - w_1^T \varphi_1(x_i) - b_1 - e_{i,1}) + \sum_{i=2}^N \alpha_{i,2}(c_{i,2}^T z - w_2^T \varphi_2(x_i) - b_2 - e_{i,2})$$

- Conditions for optimality [Fletcher, 1987]:

$$\frac{\partial \mathcal{L}}{\partial z} = 0, \frac{\partial \mathcal{L}}{\partial w_1} = 0, \frac{\partial \mathcal{L}}{\partial w_2} = 0, \frac{\partial \mathcal{L}}{\partial b_1} = 0, \frac{\partial \mathcal{L}}{\partial b_2} = 0, \frac{\partial \mathcal{L}}{\partial e_{1,1}} = 0, \frac{\partial \mathcal{L}}{\partial e_{1,2}} = 0,$$
$$\frac{\partial \mathcal{L}}{\partial \beta_{1,1}} = 0, \frac{\partial \mathcal{L}}{\partial \beta_{1,2}} = 0, \frac{\partial \mathcal{L}}{\partial \alpha_{i,1}} = 0, \frac{\partial \mathcal{L}}{\partial \alpha_{i,2}} = 0$$

Proof - conditions for optimality

$$
\begin{cases}
\frac{\partial \mathcal{L}}{\partial z} &= -\gamma z + (I - P_D)^T (I - P_D) z + \beta_{1,1} c_{1,1} + \beta_{1,2} c_{1,2} + \\
& \quad \sum_{i=2}^N \alpha_{i,1} c_{i,1} + \sum_{i=2}^N \alpha_{i,2} c_{i,2} = 0 \\
\frac{\partial \mathcal{L}}{\partial w_1} &= \nu w_1 - \sum_{i=2}^N \alpha_{i,1} \varphi_1(x_i) = 0 \\
\frac{\partial \mathcal{L}}{\partial w_2} &= \nu w_2 - \sum_{i=2}^N \alpha_{i,2} \varphi_2(x_i) = 0 \\
\frac{\partial \mathcal{L}}{\partial b_1} &= \sum_{i=2}^N \alpha_{i,1} = 1_{N-1}^T \alpha_1 = 0 \\
\frac{\partial \mathcal{L}}{\partial b_2} &= \sum_{i=2}^N \alpha_{i,2} = 1_{N-1}^T \alpha_2 = 0 \\
\frac{\partial \mathcal{L}}{\partial e_{1,1}} &= \eta e_{1,1} - \beta_{1,1} = 0 \\
\frac{\partial \mathcal{L}}{\partial e_{1,2}} &= \eta e_{1,2} - \beta_{1,2} = 0 \\
\frac{\partial \mathcal{L}}{\partial e_{i,1}} &= \eta e_{i,1} - \alpha_{i,1} = 0 \, , \; i = 2, \ldots, N \\
\frac{\partial \mathcal{L}}{\partial e_{i,2}} &= \eta e_{i,2} - \alpha_{i,2} = 0 \, , \; i = 2, \ldots, N \\
\frac{\partial \mathcal{L}}{\partial \beta_{1,1}} &= c_{1,1}^T z - q_1 - e_{1,1} = 0 \\
\frac{\partial \mathcal{L}}{\partial \beta_{1,2}} &= c_{1,2}^T z - q_2 - e_{1,2} = 0 \\
\frac{\partial \mathcal{L}}{\partial \alpha_{i,1}} &= c_{i,1}^T z - {w_1}^T \varphi_1(x_i) - b_1 - e_{i,1} = 0 \, , \; i = 2, \ldots, N \\
\frac{\partial \mathcal{L}}{\partial \alpha_{i,2}} &= c_{i,2}^T z - {w_2}^T \varphi_2(x_i) - b_2 - e_{i,2} = 0 \, , \; i = 2, \ldots, N.
\end{cases}
$$

Proof - elimination step

- Eliminate $w_1, w_2, e_{i,1}, e_{i,2}$
- Express in term of kernel functions
- Express set of equations in terms of $z, b_1, b_2, \alpha_1, \alpha_2$
- One obtains

$$-\gamma z+(I-P_D)^T(I-P_D)z+V_1\alpha_1+V_2\alpha_2+\eta c_{1,1}c_{1,1}^T z+\eta c_{1,2}c_{1,2}^T z = \eta(q_1 c_{1,1}+q_2 c_{1,2})$$

and

$$\begin{array}{l} V_1^T z - \frac{1}{\nu}\Omega_1\alpha_1 - \frac{1}{\eta}\alpha_1 - b_1 1_{N-1} = 0 \\ V_2^T z - \frac{1}{\nu}\Omega_2\alpha_2 - \frac{1}{\eta}\alpha_2 - b_2 1_{N-1} = 0 \\ \beta_{1,1} = \eta(c_{1,1}^T z - q_1) \\ \beta_{1,2} = \eta(c_{1,2}^T z - q_2). \end{array}$$

- The dual model representation follows from the conditions for optimality.

Model selection by validation

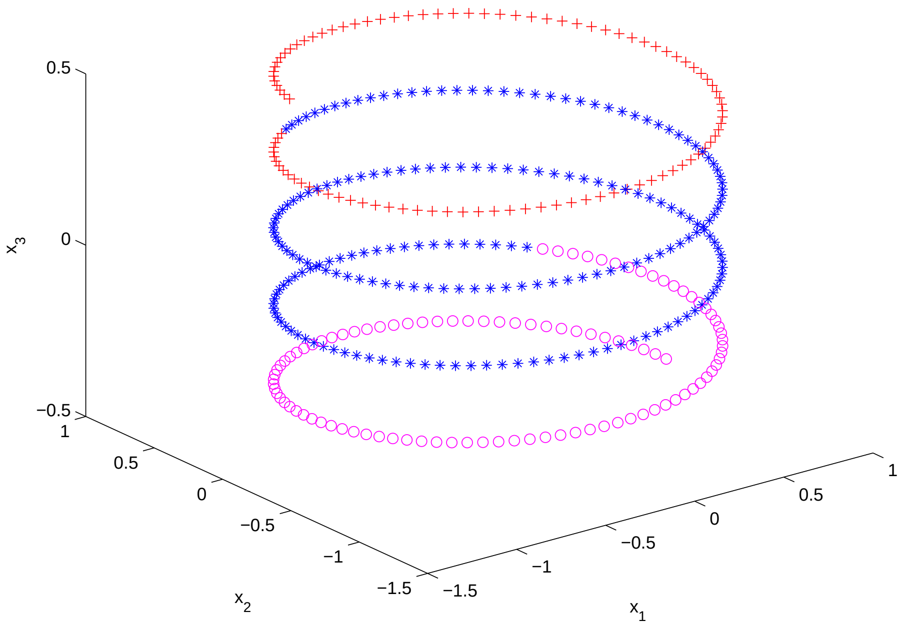
Model selection criterion:
$$\min_{\Theta} \sum_{i,j} \left(\frac{\hat{z}_i^T \hat{z}_j}{\|\hat{z}_i\|_2 \|\hat{z}_j\|_2} - \frac{x_i^T x_j}{\|x_i\|_2 \|x_j\|_2} \right)^2$$

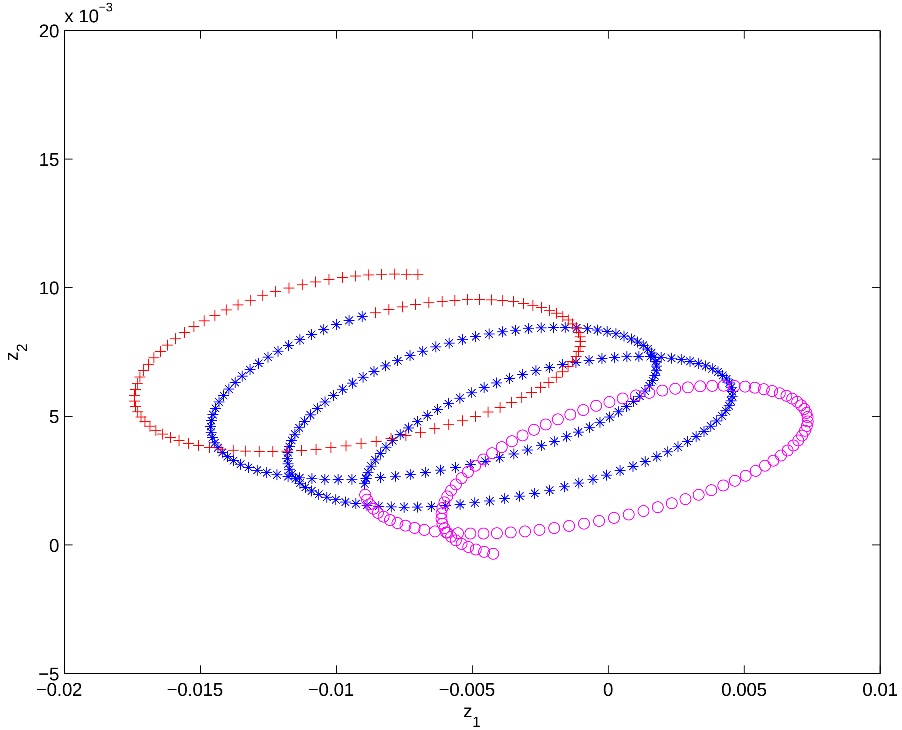
Tuning parameters Θ:

- Kernels tuning parameters in $s_{ij}, K_1, K_2, (K_3)$
- Regularization constants ν, η (take $\gamma = 0$)
- Choice of the diagonal matrix D
- Choice of reference point q, e.g. $q \in \{[+1;+1], [+1;-1], [-1;+1], [-1,-1]\}$

$\longrightarrow$ Stable results, finding a "good range" is satisfactory.

KMref: spiral example

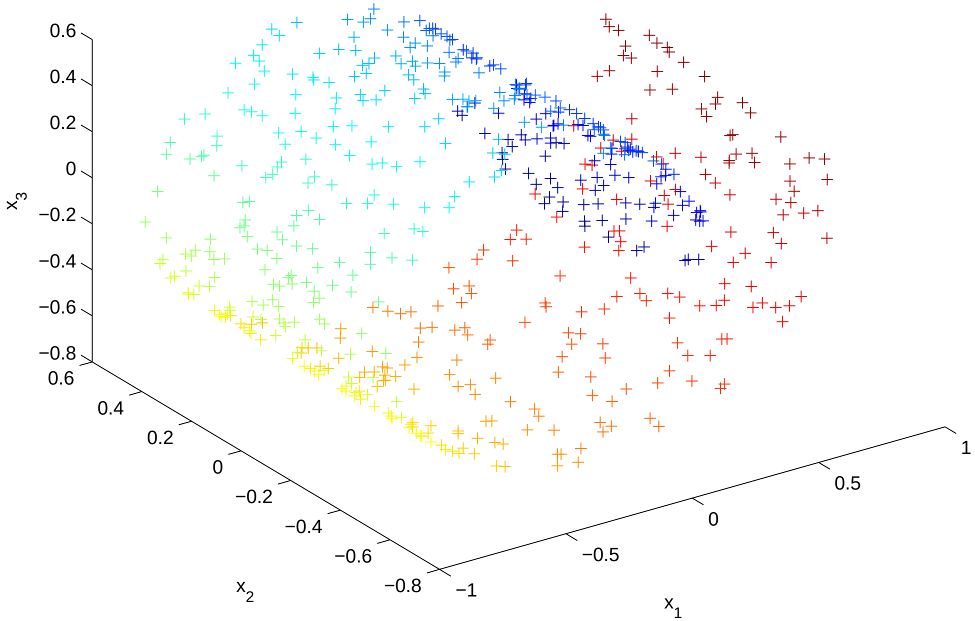




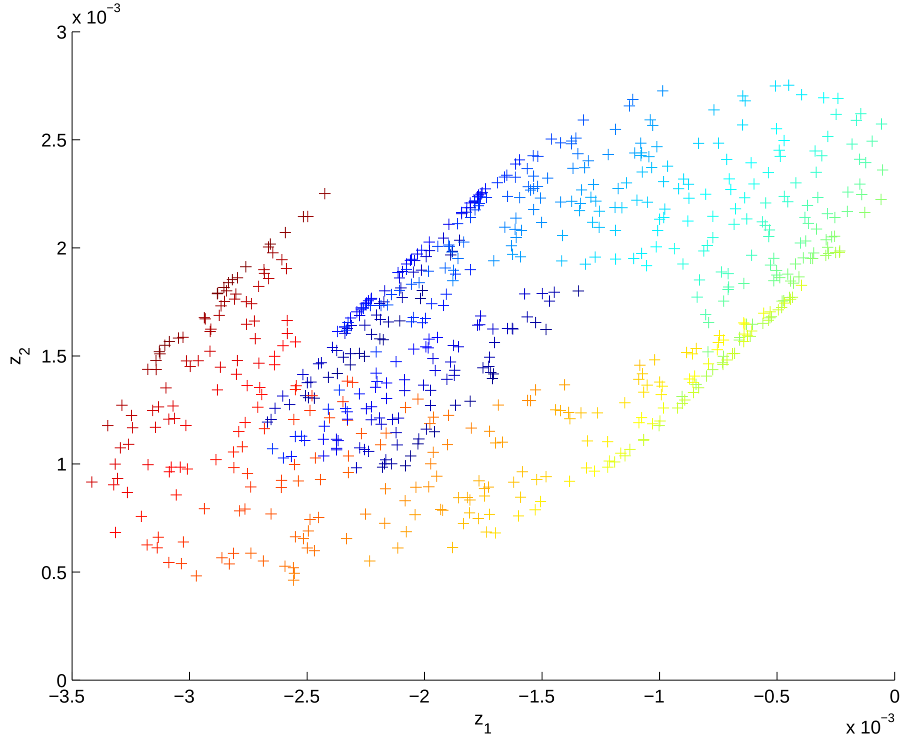
training data (blue *), validation data (magenta o), test data (red +)

Model selection: $\min \sum_{i,j} \left(\frac{\hat{z}_i^T \hat{z}_j}{\|\hat{z}_i\|_2 \|\hat{z}_j\|_2} - \frac{x_i^T x_j}{\|x_i\|_2 \|x_j\|_2} \right)^2$

KMref: swiss roll example



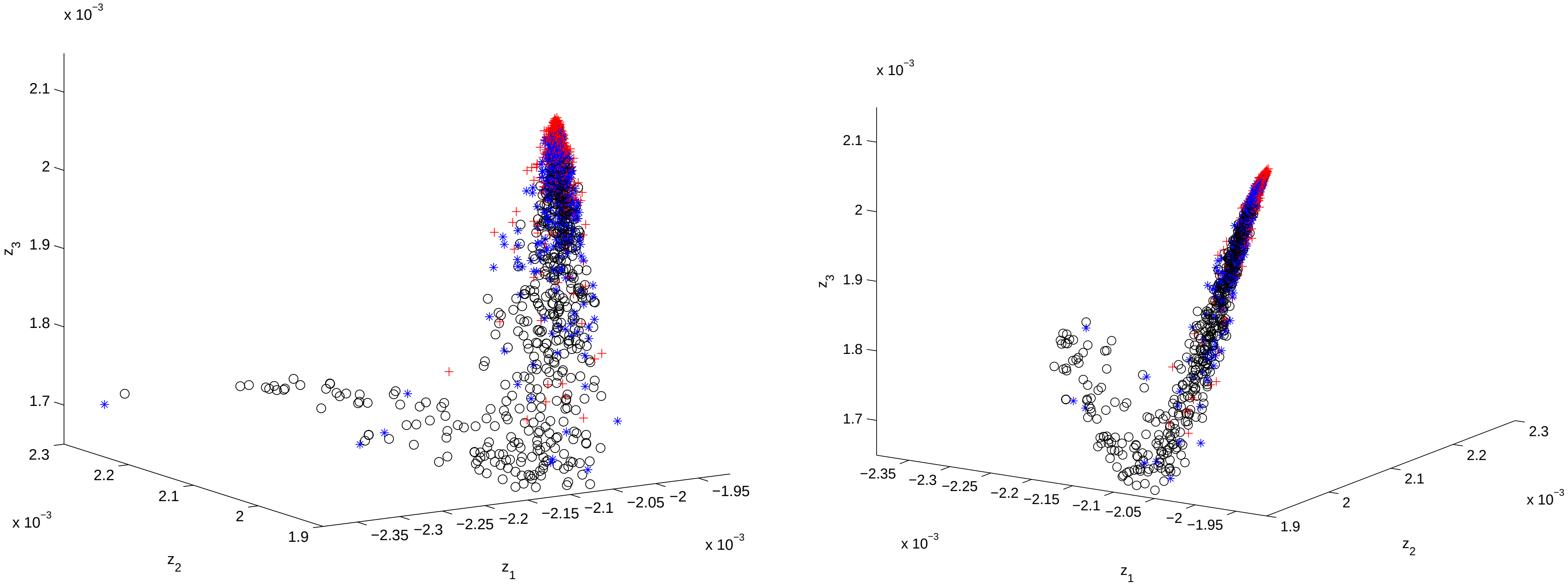
Given 3D swiss roll data



KMref result - 2D projection

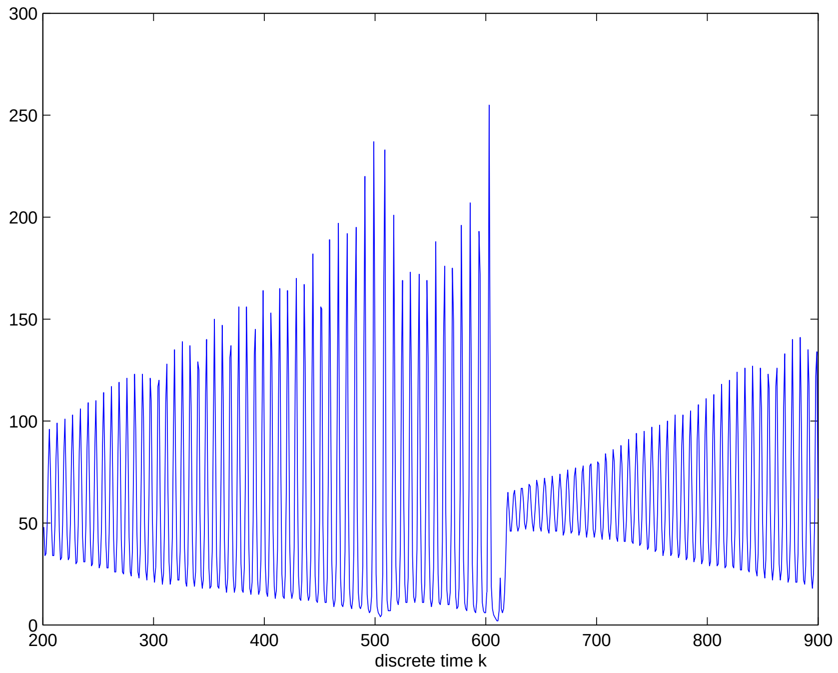
600 training data, 100 validation data

KMref: visualizing gene distributions

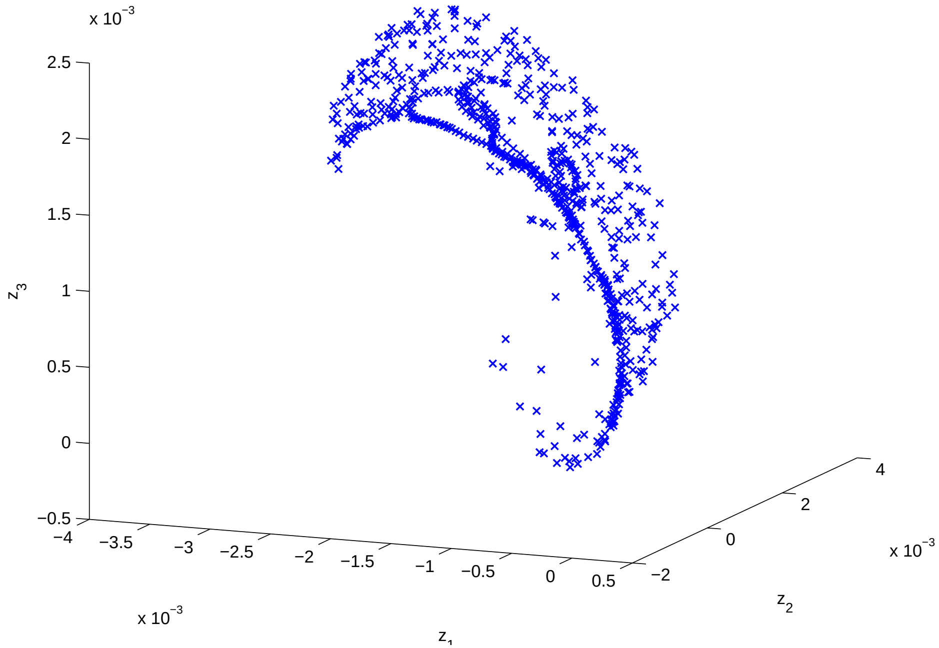


KMref 3D projection (Alon colon cancer microarray data set)
Dimension input space: 62
Number of genes: 1500 (training: 500, validation: 500, test: 500)
Model selection: $\sigma^2 = 10^4$, $\sigma_1^2 = 10^3$, $\sigma_2^2 = 0.5\sigma_1^2$, $\sigma_3^2 = 0.1\sigma_1^2$,
$\eta = 1$, $\nu = 100$, $D = \mathrm{diag}\{10, 5, 1\}$, $q = [+1; -1; -1]$.

KMref: Santa Fe laser data



original time-series $\{y_t\}_{t=1}^{t=T}$



3D projection

construct $y_{t|t-m} = [y_t; y_{t-1}; y_{t-2}; ...; y_{t-m}]$ with $m = 9$
given data $\{y_{t|t-m}\}_{t=m+1}^{t=m+N_{tot}}$ in a $p = 10$ dimensional space
200 validation data (first part), 700 training data points

Conclusions

- Trend: "Kernelizing" classical methods (FDA, PCA, CCA, ICA, ...)
- Kernel methods: complementary views (LS-)SVM, RKHS, GP
- Least squares support vector machines as "core problems" in supervised and unsupervised learning, and beyond
- LS-SVM provides methodology for *"optimization modelling"*
  Kernel maps with reference point: LS-SVM core part
  Computational complexity: similar to regression/classification
- Reference point: converts eigenvalue problem into linear system

Read more: http://www.esat.kuleuven.be/sista/lssvmlab/KMref/KMref0722.pdf

Matlab demo file: http://www.esat.kuleuven.be/sista/lssvmlab/KMref/demoswissKMref.m