Challenges of Virtual Testing in Statistical Quality Control of Railway Ballast

Vera Hofer

University of Graz

Ljubljana, 21/04/2015

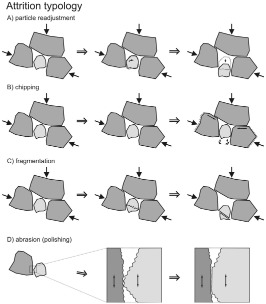
Why SPC of railway ballast?

Source: www.erdwissen.ch

- Technical aspect
  - Forces act on the track bed
  - Attrition

Why SPC of railway ballast?

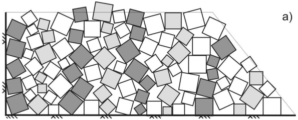
Source: www.erdwissen.ch



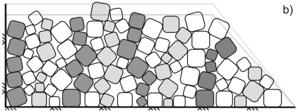
- Technical aspect
  - Forces act on the track bed
  - Attrition
  - Operating risk!

Why SPC of railway ballast?

Source: www.erdwissen.ch



perfectly angular particles

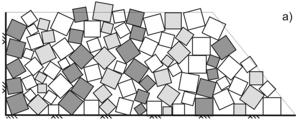


rounded particles
- Volume loss of particles 5,4%
- Volume loss of heap 10,5%

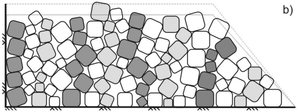
- Technical aspect
  - Forces act on the track bed
  - Attrition
  - Operating risk!
- Economic aspect
  - Life cycle costs determined by cost of maintenance
  - Maintenance operations: temporary track closure and downtime costs

Why SPC of railway ballast?

Source: www.erdwissen.ch



perfectly angular particles

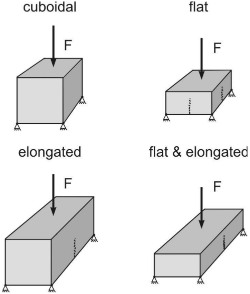


rounded particles
- Volume loss of particles 5,4%
- Volume loss of heap 10,5%

- Technical aspect
  - Forces act on the track bed
  - Attrition
  - Operating risk!
- Economic aspect
  - Life cycle costs determined by cost of maintenance
  - Maintenance operations: temporary track closure and downtime costs
- Quality of the material
  - Geometry
  - Rock type

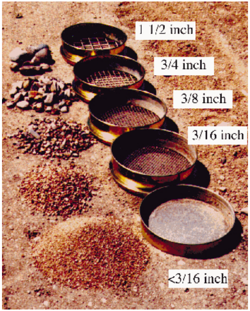
Why SPC of railway ballast?

Source: www.erdwissen.ch



- Technical aspect
  - Forces act on the track bed
  - Attrition
  - Operating risk!
- Economic aspect
  - Life cycle costs determined by cost of maintenance
  - Maintenance operations: temporary track closure and downtime costs
- Quality of the material
  - Geometry
  - Rock type

Why SPC of railway ballast?



Problem of determining geometry, rock type, mechanical properties (abrasion, fragmentation, resistence to wear)

- Manual tests time consuming and expensive
- Laboratory for mechanical properties (LA value, other measures)

Why SPC of railway ballast?

www.pavementinteractive.org

Problem of determining geometry, rock type, mechanical properties (abrasion, fragmentation, resistence to wear)

- Manual tests time consuming and expensive
- Laboratory for mechanical properties (LA value, other measures)

Why SPC of railway ballast?

www.pavementinteractive.org

Problem of determining geometry, rock type, mechanical properties (abrasion, fragmentation, resistence to wear)

- Manual tests time consuming and expensive
- Laboratory for mechanical properties (LA value, other measures)

Objective: Replace manual tests by statistical prediction (= “virtual testing”)

Virtual testing

www.pavementinteractive.org

Major task:

- Prediction of mechanical properties Y (for example LA value)
  - Rock type X_1
  - Shape X_2
  - Size X_3
  - Angularity X_4

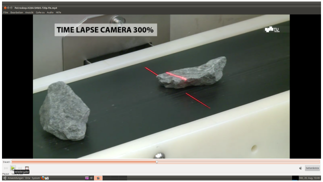
Virtual testing

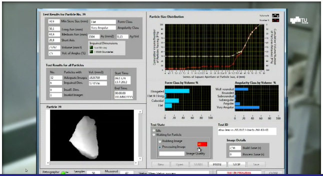
www.pavementinteractive.org

Major task:

- Prediction of mechanical properties Y (for example LA value)
  - Rock type X_1
  - Shape X_2
  - Size X_3
  - Angularity X_4

Virtual testing

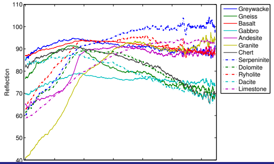




Major task:

- Prediction of mechanical properties Y (for example LA value)
  - Rock type X_1
  - Shape X_2
  - Size X_3
  - Angularity X_4

Virtual testing

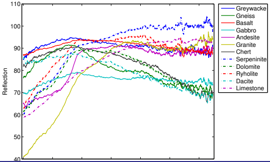


Major task:

- Prediction of mechanical properties Y (for example LA value)
  - Rock type X_1
  - Shape X_2
  - Size X_3
  - Angularity X_4

- Special issues of measurement
  - Geometric and spectroscopic features (predictors) are measured particlewise, not linked
  - Mechanical properties Y are measured samplewise (several hundered particles

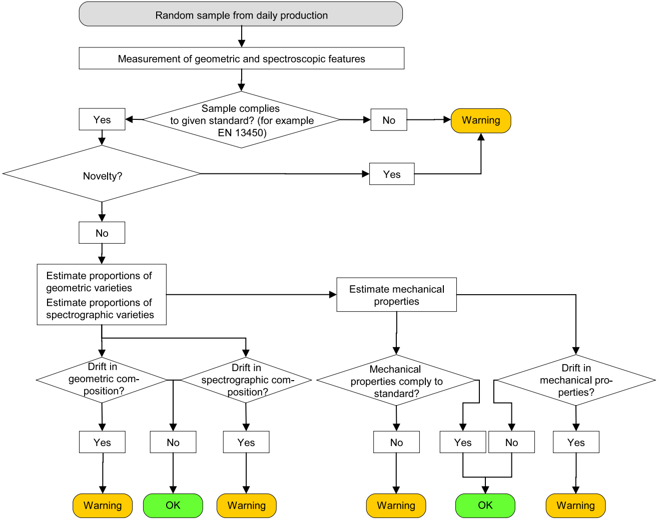
Virtual testing



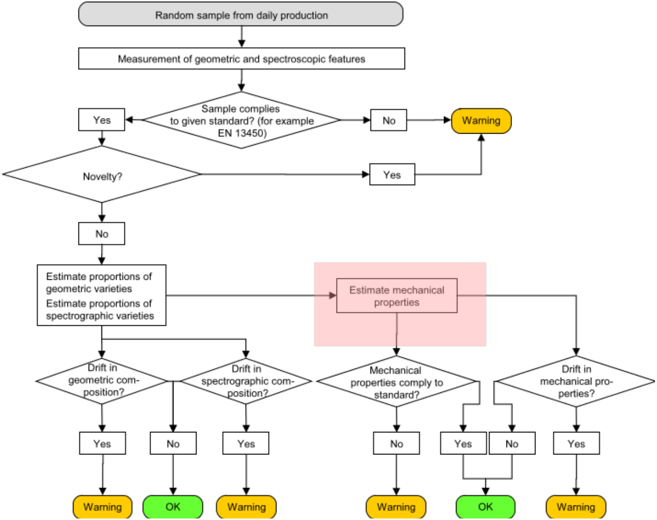
Major task:

- Prediction of mechanical properties Y (for example LA value)
  - Rock type X_1
  - Shape X_2
  - Size X_3
  - Angularity X_4
- Special issues of measurement
  - Geometric and spectroscopic features (predictors) are measured particlewise, not linked
  - Mechanical properties Y are measured samplewise (several hundered particles

Statistical monitoring system

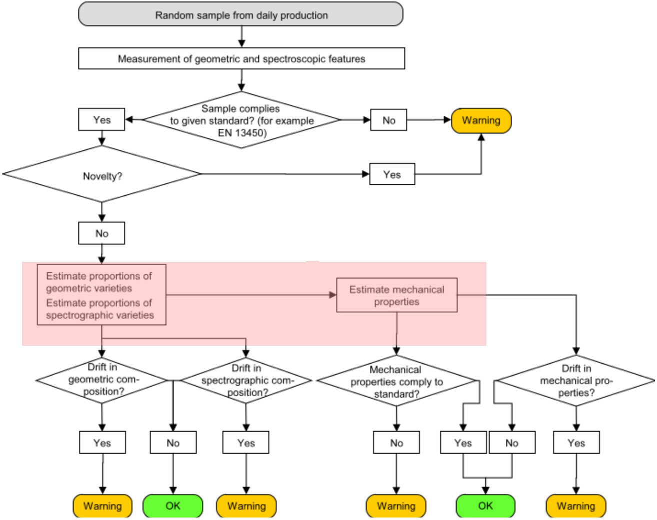


Statistical monitoring system



Statistical monitoring system

Statistical monitoring system



Compositional data approach

- Aggregation of particlewise geometric property X_i for sample $j = 1, 2, \ldots$ to the samplewise feature $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iL})$ with

$$\sum_{j=1}^{L} p_{ij} = 1$$

- Raw data are continuous features: no direct aggregation!
  - Cluster analysis to determine geometric similarity classes and classification of new sample into these classes
  - Classification of spectroscopic data into rock classes
  - Determine distribution vectors $\mathbf{q}_i$ for rock composition and $\mathbf{p}_i$ for geometric composition

Compositional data approach

- Aggregation of particlewise geometric property X_i for sample $j = 1, 2, \ldots$ to the samplewise feature $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iL})$ with

$$\sum_{j=1}^{L} p_{ij} = 1$$

- Raw data are continuous features: no direct aggregation!
  - Cluster analysis to determine geometric similarity classes and classification of new sample into these classes
  - Classification of spectroscopic data into rock classes
  - Determine distribution vectors $\mathbf{q}_i$ for rock composition and $\mathbf{p}_i$ for geometric composition
- Predictors (and possibly response variable) are compositional vectors

Compositional data approach

- Aggregation of particlewise geometric property X_i for sample $j = 1, 2, \ldots$ to the samplewise feature $\mathbf{p}_i = (p_{i1}, p_{i2}, \ldots, p_{iL})$ with

$$\sum_{j=1}^{L} p_{ij} = 1$$

- Raw data are continuous features: no direct aggregation!
  - Cluster analysis to determine geometric similarity classes and classification of new sample into these classes
  - Classification of spectroscopic data into rock classes
  - Determine distribution vectors $\mathbf{q}_i$ for rock composition and $\mathbf{p}_i$ for geometric composition
- Predictors (and possibly response variable) are compositional vectors

Compositional data versus multivariate data

- Negative bias: For $X = (X_1, \ldots, X_L)$ with $\sum_{i=1}^{L} X_i = 1$

$$
\begin{aligned}
Cov(X_1, \sum_{i=1}^{L} X_i) &= Cov(X_1, 1) = 0 \\
\sum_{i=2}^{L} Cov(X_1, X_i) &= -Var(X_1)
\end{aligned}
$$

- Spurious correlation

Compositional data versus multivariate data

- Negative bias
- Spurious correlation: correlation of the ratios X/Z and Y/Z of three uncorrelated random variables will not be uncorrelated (Pearson, 1897)
- Euclidean distance can be larger in a subspace than for the full composition

Compositional data versus multivariate data

- Negative bias
- Spurious correlation
- Euclidean distance can be larger in a subspace than for the full composition
- Variance matrices are singular due to the constant-sum constraint

Compositional data versus multivariate data

- Negative bias
- Spurious correlation
- Euclidean distance can be larger in a subspace than for the full composition
- Variance matrices are singular due to the constant-sum constraint
- Univariate analyis makes no sense, since the value of one component only meaningful in relation to the other components

Compositional data versus multivariate data

- Negative bias
- Spurious correlation
- Euclidean distance can be larger in a subspace than for the full composition
- Variance matrices are singular due to the constant-sum constraint
- Univariate analyis makes no sense, since the value of one component only meaningful in relation to the other components

Compositional data versus multivariate data

- Negative bias
- Spurious correlation
- Euclidean distance can be larger in a subspace than for the full composition
- Variance matrices are singular due to the constant-sum constraint
- Univariate analyis makes no sense, since the value of one component only meaningful in relation to the other components

Compositional data analysis!

Pioneer

Prof. Aitchison

Source: Wikipedia

Isometric log-ratio (ilr) transform

- Let **p** be a compositional data vector in the simplex

$$\mathbb{S}^L = \left\{ \mathbf{p} = (p_1, \ldots, p_L) \in [0,\, 1]^L \;\middle|\; \sum_{i=1}^{L} p_i = 1 \right\}$$

- ilr transformation: $ilr : \mathbb{S}^L \longrightarrow \mathbb{R}^{L-1}$

$$\mathbf{x} = \mathrm{ilr}(\mathbf{p}) = \ln(\mathbf{p})\,\Psi,$$

Isometric log-ratio (ilr) transform

- Let $\mathbf{p}$ be a compositional data vector in the simplex

$$\mathbb{S}^L = \left\{ \mathbf{p} = (p_1, \dots, p_L) \in [0,\,1]^L \;\middle|\; \sum_{i=1}^{L} p_i = 1 \right\}$$

- ilr transformation: $ilr : \mathbb{S}^L \longrightarrow \mathbb{R}^{L-1}$

$$\mathbf{x} = \mathrm{ilr}(\mathbf{p}) = \ln(\mathbf{p})\,\boldsymbol{\Psi},$$

where $\boldsymbol{\Psi}$ is the $L \times (L-1)$ Helmert matrix with normalised columns

$$\boldsymbol{\Psi} = \begin{pmatrix} \frac{L-1}{\sqrt{L(L-1)}} & 0 & \cdots & 0 & 0 \\ -\frac{1}{\sqrt{L(L-1)}} & \frac{L-2}{\sqrt{(L-1)(L-2)}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{\sqrt{L(L-1)}} & \frac{-1}{\sqrt{(L-1)(L-2)}} & \cdots & \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \end{pmatrix}$$

- Now work with coordinates $\mathbf{x} \in \mathbb{R}^{L-1}$

Isometric log-ratio (ilr) transform

- Let **p** be a compositional data vector in the simplex

$$\mathbb{S}^L = \left\{ \mathbf{p} = (p_1, \ldots, p_L) \in [0,\, 1]^L \;\middle|\; \sum_{i=1}^{L} p_i = 1 \right\}$$

- ilr transformation: $ilr : \mathbb{S}^L \longrightarrow \mathbb{R}^{L-1}$

$$\mathbf{x} = \mathrm{ilr}(\mathbf{p}) = \ln(\mathbf{p})\, \boldsymbol{\Psi},$$

- Now work with coordinates $\mathbf{x} \in \mathbb{R}^{L-1}$
- Inverse transform ilr^{-1}

$$ilr^{-1}(\mathbf{x}) = \mathcal{C}(exp(\mathbf{x} \cdot \boldsymbol{\Psi}')),$$

where $\mathcal{C}$ is the closure operations, i.e. $\mathcal{C}(\mathbf{y}) = \frac{1}{\sum_i y_i}\mathbf{y}$

Isometric log-ratio (ilr) transform

- Let **p** be a compositional data vector in the simplex

$$\mathbb{S}^L = \left\{ \mathbf{p} = (p_1, \ldots, p_L) \in [0, 1]^L \;\middle|\; \sum_{i=1}^{L} p_i = 1 \right\}$$

- ilr transformation: $ilr : \mathbb{S}^L \longrightarrow \mathbb{R}^{L-1}$

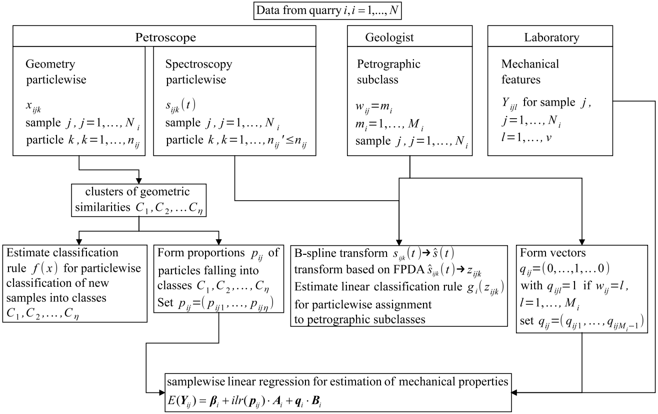
$$\mathbf{x} = \mathrm{ilr}(\mathbf{p}) = \ln(\mathbf{p})\, \mathbf{\Psi},$$

- Now work with coordinates $\mathbf{x} \in \mathbb{R}^{L-1}$
- Inverse transform ilr^{-1}

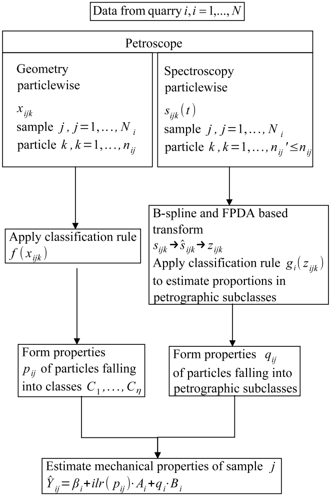
$$ilr^{-1}(\mathbf{x}) = \mathcal{C}(exp(\mathbf{x} \cdot \mathbf{\Psi}')),$$

where $\mathcal{C}$ is the closure operations, i.e. $\mathcal{C}(\mathbf{y}) = \frac{1}{\sum_i y_i}\, \mathbf{y}$

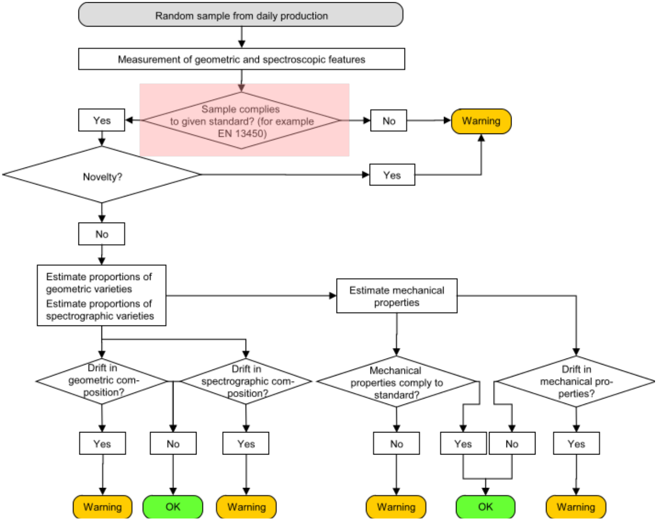
Statistical modelling



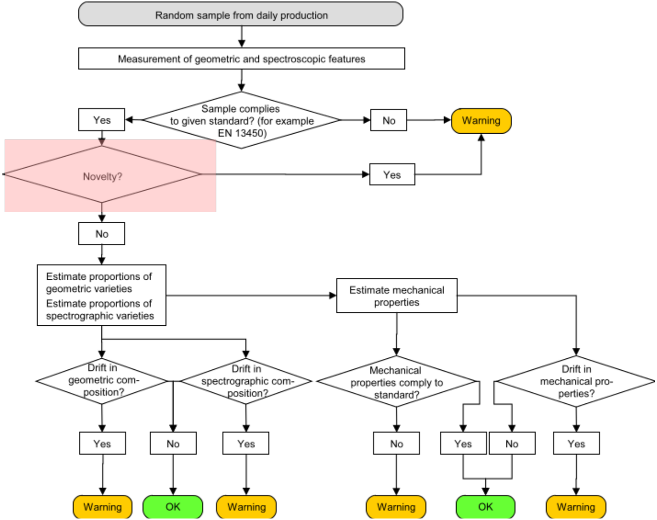
Statistical prediction



Statistical monitoring system



Statistical monitoring system



Novelty detection

- New spectrographic varieties $\longrightarrow$ wrong prediction of spectroscopic composition and of mechanical properties
- Novelty detection by means of one-class support vector machines:
  - Extract a subset of 'regular' data points from a given random dataset

Novelty detection

- New spectrographic varieties $\longrightarrow$ wrong prediction of spectroscopic composition and of mechanical properties
- Novelty detection by means of one-class support vector machines:
  - Extract a subset of 'regular' data points from a given random dataset
  - Dataset contains reflectance spectra

$$s_{ijk_2}(t) \approx \hat{s}_{ijk_2}(t) = \sum_{l=1}^{D} \lambda_{ijk_2 l}\, u_l(t) = \boldsymbol{\lambda}'_{ijk_2}\, \mathbf{u}(t)\,,$$

Novelty detection

- New spectrographic varieties $\longrightarrow$ wrong prediction of spectroscopic composition and of mechanical properties
- Novelty detection by means of one-class support vector machines:
  - Extract a subset of ‘regular’ data points from a given random dataset
  - Dataset contains reflectance spectra

$$s_{ijk_2}(t) \approx \hat{s}_{ijk_2}(t) = \sum_{l=1}^{D} \lambda_{ijk_2 l}\, u_l(t) = \boldsymbol{\lambda}'_{ijk_2}\, \mathbf{u}(t)\,,$$

  - PC transformation of the basis coefficients $\mathbf{z}'_{ijk_2} = (z_{ijk_2 1}, \ldots, z_{ijk_2 E})$

Novelty detection

- New spectrographic varieties $\longrightarrow$ wrong prediction of spectroscopic composition and of mechanical properties
- Novelty detection by means of one-class support vector machines:
  - Extract a subset of 'regular' data points from a given random dataset
  - Dataset contains reflectance spectra

$$s_{ijk_2}(t) \approx \hat{s}_{ijk_2}(t) = \sum_{l=1}^{D} \lambda_{ijk_2 l}\, u_l(t) = \boldsymbol{\lambda}'_{ijk_2}\, \mathbf{u}(t)\,,$$

  - PC transformation of the basis coefficients $\mathbf{z}'_{ijk_2} = (z_{ijk_2 1}, \ldots, z_{ijk_2 E})$

One-class support vector machines

- Decision rule

$$g_i(\mathbf{z}) = \mathrm{sign}\left(\sum_{j=1}^{N_i}\sum_{k_2=1}^{n'_{ij}} \theta_{ijk_2} K(\mathbf{z}, \mathbf{z}_{ijk_2}) - \theta_0\right)$$

with $g_i(\mathbf{z}) = +1$ for ‘regular’ data and $g_i(\mathbf{z}) = -1$ for ‘novel’ data is estimated.

- $K(\cdot,\cdot)$ denotes the radial kernel

$$K(\mathbf{z}, \mathbf{z}_{ijk_2}) = e^{\delta \, \|\mathbf{z}-\mathbf{z}_{ijk_2}\|^2}$$

with parameter δ.

One-class support vector machines

Solve the minimisation problem

$$\min_{\boldsymbol{\theta}_\mathbf{i}} \quad \frac{1}{2} \sum_{j,j'=1}^{N_i} \sum_{k_2,k_2'=1}^{n'_{ij}} \theta_{ijk_2} \theta_{ij'k_2'} K(\mathbf{z}_{ijk_2}, \mathbf{z}_{ij'k_2'})$$

$$\text{s.t.} \quad \sum_{j=1}^{N_i} \sum_{k_2=1}^{n'_{ij}} \theta_{ijk_2} = 1$$

$$0 \leq \theta_{ijk_2} \leq \frac{1}{\rho\, n'_i} \qquad n'_i = \sum_{j=1}^{N_i} n'_{ij}$$

One-class support vector machines

Solve the minimisation problem

$$\min_{\boldsymbol{\theta}_\mathbf{i}} \quad \frac{1}{2}\sum_{j,j'=1}^{N_i}\sum_{k_2,k_2'=1}^{n'_{ij}} \theta_{ijk_2}\theta_{ij'k_2'}K(\mathbf{z}_{ijk_2},\mathbf{z}_{ij'k_2'})$$

$$\text{s.t.} \quad \sum_{j=1}^{N_i}\sum_{k_2=1}^{n'_{ij}} \theta_{ijk_2} = 1$$

$$0 \leq \theta_{ijk_2} \leq \frac{1}{\rho\, n'_i} \qquad n'_i = \sum_{j=1}^{N_i} n'_{ij}$$

$\rho \in]0,\, 1]$ is an upper bound on the number of training points outside the location of 'regular' data points.

One-class support vector machines

Solve the minimisation problem

$$\min_{\boldsymbol{\theta}_i} \quad \frac{1}{2} \sum_{j,j'=1}^{N_i} \sum_{k_2,k_2'=1}^{n'_{ij}} \theta_{ijk_2}\theta_{ij'k_2'} K(\mathbf{z}_{ijk_2}, \mathbf{z}_{ij'k_2'})$$

$$\text{s.t.} \quad \sum_{j=1}^{N_i} \sum_{k_2=1}^{n'_{ij}} \theta_{ijk_2} = 1$$

$$0 \leq \theta_{ijk_2} \leq \frac{1}{\rho\, n'_i} \qquad n'_i = \sum_{j=1}^{N_i} n'_{ij}$$

θ_0 is derived for a point with $0 < \theta_{ijk_2} < \frac{1}{\rho\, n'_i}$ as

$$\theta_0 = \sum_{j'=1}^{N_i} \sum_{k_2'=1}^{n'_{ij}} \theta_{ij'k_2'} K(\mathbf{z}_{ij'k_2'}, \mathbf{z}_{ijk_2}) \, .$$

Statistical monitoring system