Proceedings of Machine Learning Research 91:1–20, 2018 Conformal and Probabilistic Prediction and Applications

Conformal Prediction in Manifold Learning

Alexander Kuleshov A.KULESHOV@SKOLTECH.RU
Skolkovo Institute of Science and Technology (Moscow, Russia)

Alexander Bernstein A.BERNSTEIN@SKOLTECH.RU
Skolkovo Institute of Science and Technology (Moscow, Russia)
Institute for Information Transmission Problems (Moscow, Russia)

Evgeny Burnaev E.BURNAEV@SKOLTECH.RU
Skolkovo Institute of Science and Technology (Moscow, Russia)
Institute for Information Transmission Problems (Moscow, Russia)

Editor: Alex Gammerman, Vladimir Vovk, Zhiyuan Luo, Evgueni Smirnov and Ralf Peeters

Abstract

The paper presents a geometrically motivated view on conformal prediction applied to nonlinear multi-output regression tasks for obtaining valid measure of accuracy of Manifold Learning Regression algorithms. A considered regression task is to estimate an unknown smooth mapping $\mathbf{f}$ from q-dimensional inputs $\mathbf{x} \in \mathbf{X}$ to m-dimensional outputs $\mathbf{y} = \mathbf{f}(\mathbf{x})$ based on training dataset $\mathbf{Z}_{(n)}$ consisting of "input-output" pairs $\{Z_i = (\mathbf{x}_i, \mathbf{y}_i = \mathbf{f}(\mathbf{x}_i))^{\mathrm{T}}, i = 1, 2, \ldots, n\}$. Manifold Learning Regression (MLR) algorithm solves this task using Manifold learning technique. At first, unknown q-dimensional Regression manifold $\mathbf{M}(\mathbf{f}) = \{(\mathbf{x}, \mathbf{f}(\mathbf{x}))^{\mathrm{T}} \in \mathbb{R}^{q+m} : \mathbf{x} \in \mathbf{X} \subset \mathbb{R}^q\}$, embedded in ambient $(q+m)$-dimensional space, is estimated from the training data $\mathbf{Z}_{(n)}$, sampled from this manifold. The constructed estimator $\mathbf{M}_{MLR}$, which is also q-dimensional manifold embedded in ambient space $\mathbb{R}^{q+m}$, is close to $\mathbf{M}$ in terms of Hausdorff distance. After that, an estimator $\mathbf{f}_{MLR}$ of the unknown function $\mathbf{f}$, mapping arbitrary input $\mathbf{x} \in \mathbf{X}$ to output $\mathbf{f}_{MLR}(\mathbf{x})$, is constructed as the solution to the equation $\mathbf{M}(\mathbf{f}_{MLR}) = \mathbf{M}_{MLR}$. Conformal prediction allows constructing a prediction region for an unknown output $\mathbf{y} = \mathbf{f}(\mathbf{x})$ at Out-of-Sample input point $\mathbf{x}$ for a given confidence level using given nonconformity measure, characterizing to which extent an example $Z = (\mathbf{x}, \mathbf{y})^{\mathrm{T}}$ is different from examples in the known dataset $\mathbf{Z}_{(n)}$. The paper proposes a new nonconformity measure based on MLR estimators using an analog of Bregman distance.

Keywords: Nonlinear Multi-Output Regression, Conformal Prediction, Manifold Learning, Manifold Learning Regression

1. Introduction

1.1. Regression Problem

The general goal of Statistical Learning/Regression is to find a data-based predictive function Vapnik (1998); James et al. (2014); Hastie et al. (2009). We consider a nonlinear multi-output regression task. Let $\mathbf{f}$ be an unknown smooth mapping from an Input space $\mathbf{X} \subset \mathbb{R}^q$ to m-dimensional Euclidean space $\mathbb{R}^m$. Given a training data set

$$\mathbf{x} \in \mathbf{Z}_{(n)} = \left\{ Z_i = \begin{pmatrix} \mathbf{x}_i \\ \mathbf{y} = \mathbf{f}(\mathbf{x}_i) \end{pmatrix}, i = 1, 2, \ldots, n \right\}, \tag{1}$$

consisting of the input-output pairs, the task is to construct function $\mathbf{f}^*$, which predicts output $\mathbf{y}^* = \mathbf{f}^*(\mathbf{x}) = \mathbf{f}^*(\mathbf{x}|\mathbf{Z}_{(n)})$ for arbitrary *previously unseen* Out-of-Sample (OoS) input $\mathbf{x} \in \mathbf{X}$ with sufficiently small predictive error

$$\delta(\mathbf{x}) = \delta(\mathbf{x}|\mathbf{y}) = |\mathbf{y}^* - \mathbf{y}| = |\mathbf{f}^*(\mathbf{x}) - \mathbf{f}(\mathbf{x})| \tag{2}$$

for a *true* output $\mathbf{y} = \mathbf{f}(\mathbf{x})$.

There are various models and methods for reconstruction of an unknown function using sample data, such as least squares techniques (linear and nonlinear), artificial neural networks, kernel nonparametric regression and kriging, SVM-regression, Gaussian processes, Radial Basic Functions, Deep Learning Networks, Gradient Boosting Regression Trees, Random Forests, etc. Hastie et al. (2009); Bishop (2006); Deng and Yu (2014); Breiman (2001); Friedman (2000); Rasmussen and Williams (2005); Belyaev et al. (2015b); Burnaev and Vovk (2014); Burnaev et al. (2016); Burnaev and Nazarov (2016); Burnaev and Panov (2015); Loader (1999); Belyaev et al. (2016, 2015a); Burnaev et al. (2013); Zaitsev et al. (2013).

Sometimes, in the framework of the regression task an estimation problem of an unknown $m \times q$ Jacobian matrix $\mathbf{J}_{\mathbf{f}}(\mathbf{x}) = \nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x})$ of the mapping $\mathbf{f}(\mathbf{x})$ at arbitrary input point $\mathbf{x} \in \mathbf{X}$ is considered.

1.2. Manifold Learning Regression

Many regression methods (kernel nonparametric regression, kriging, Gaussian processes) use kernels that are typically stationary. However, as indicated in many studies Hastie et al. (2009); Rasmussen and Williams (2005); Loader (1999), application of these methods result in significant prediction errors in case functions with strongly varying gradients; thus, it is necessary to use non-stationary kernels with adaptive kernel width.

To avoid this drawback we proposed in Cayton (2005); Huo et al. (2011); Ma and Fu (2011) a new geometrically motivated Manifold Learning Regression (MLR) algorithm based on Manifold Learning approach. Let us describes its main idea. An unknown smooth function $\mathbf{f}$ determines an unknown smooth q-dimensional manifold

$$\mathbf{M}(\mathbf{f}) = \left\{ Z = \mathbf{F}(\mathbf{x}) = \begin{pmatrix} \mathbf{x} \\ \mathbf{f}(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^{q+m} : \mathbf{x} \in \mathbf{X} \subset \mathbb{R}^q \right\}, \tag{3}$$

which is called Regression manifold (RM), embedded in an ambient space $\mathbb{R}^p$, $p = q+m$, and covered (parameterized) by a single chart $\mathbf{F} : \mathbf{x} \in \mathbf{X} \subset \mathbb{R}^q \to Z = \mathbf{F}(\mathbf{x}) \in \mathbf{M}(\mathbf{f}) \subset \mathbb{R}^p$. Any estimated function $\mathbf{f}^* : \mathbf{X} \to \mathbb{R}^m$ also determines a manifold $\mathbf{M}(\mathbf{f}^*)$ by replacing function $\mathbf{f}(\mathbf{x})$ with $\mathbf{f}^*(\mathbf{x})$ in (3).

At first, the MLR estimates an unknown Regression manifold $\mathbf{M}(\mathbf{f})$ (3) using training data $\mathbf{Z}_{(n)}$, which can be considered as a sample from this manifold. The constructed estimator $\mathbf{M}_{MLR} = \mathbf{M}_{MLR}(\mathbf{Z}_{(n)})$, which is also a q-dimensional manifold embedded in ambient space $\mathbb{R}^{q+m}$, provides a small Hausdorff distance $d_H(\mathbf{M}_{MLR}, \mathbf{M}(\mathbf{f}))$ between these manifolds. After that, an estimator $\mathbf{f}_{MLR}$ of the unknown function $\mathbf{f}$ is constructed as the solution of the equation

$$\mathbf{M}(\mathbf{f}_{MLR}) = \mathbf{M}_{MLR}(\mathbf{Z}_{(n)}).$$

The MLR also estimates an unknown $m \times q$ Jacobian matrix $\mathbf{J}_{\mathbf{f}}(\mathbf{x}) = \nabla_{\mathbf{x}}\mathbf{f}(\mathbf{x})$ of the mapping $\mathbf{f}(\mathbf{x})$ by $m \times q$ matrix $\mathbf{G}_{MLR}(\mathbf{x}) = \mathbf{G}_{MLR}(\mathbf{x}|\mathbf{Z}_{(n)})$ at arbitrary input point $\mathbf{x} \in \mathbf{X}$.

Under appropriate choice of parameters in MLR procedures, the errors of MLR-estimators $\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{f}_{MLR}(\mathbf{x})$ and $\mathbf{G}_{MLR}(\mathbf{x})$ have the following convergence rates:

$$|\hat{\mathbf{y}}(\mathbf{x}) - \mathbf{f}(\mathbf{x})| \sim O(n^{-2/(q+2)}), \tag{4}$$

$$|\mathbf{G}_{MLR}(\mathbf{x}) - \mathbf{J}_f(\mathbf{x})| = O(n^{-1/(q+2)}). \tag{5}$$

Kuleshov and Bernstein (2016, 2017) generalize the manifold regression technique to the case in which the Input space $\mathbf{X} \subset \mathbb{R}^q$ is an unknown Input manifold with known lower intrinsic dimensionality $q' < q$, embedded in ambient space $\mathbb{R}^q$.

1.3. Conformal prediction

Conformal prediction technique Vovk et al. (2005); Shafer and Vovk (2008); Papadopoulos et al. (2014) is used for determining the precise level of confidence in prediction $\mathbf{f}_{MLR}(\mathbf{x})$ for arbitrary input $\mathbf{x} \in \mathbf{X}$ by constructing the prediction region $\Gamma_\alpha(\mathbf{x}) = \Gamma_\alpha(\mathbf{x}|\mathbf{Z}_{(n)})$ which contains the unknown value $\mathbf{y} = \mathbf{f}(\mathbf{x})$ with a given confidence level α.

Common approach to the Conformal prediction is as follows Shafer and Vovk (2008). At first, a non-conformity measure is introduced as a real-valued function $A(\mathbf{Z}_{(n)}; Z)$ that measures how different an example $Z = (\mathbf{X}, \mathbf{y})^{\mathrm{T}}$ from the examples in known dataset $\mathbf{Z}_{(n)}$. Note that if input $\mathbf{x}$ doesn't belong to the sample dataset of inputs

$$\mathbf{X}_{(n)} = \{\mathbf{x}_i, i = 1, 2, \ldots, n\}, \tag{6}$$

the output $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is unknown, and the nonconformity measure A can be written only as a real-valued function

$$a(\mathbf{y}) = a(\mathbf{y}|\mathbf{x}, \mathbf{Z}_{(n)}) = A(\mathbf{Z}_{(n)}; (\mathbf{x}, \mathbf{y})^{\mathrm{T}})$$

of output $\mathbf{y}$ given dataset $\mathbf{X}_{(n)}$ and input $\mathbf{x}$.

Given some estimating procedure $\mathbf{f}^*$, which determines predicted value

$$\mathbf{y}^* = \mathbf{f}^*(\mathbf{x}) = \mathbf{f}^*(\mathbf{x}|\mathbf{Z}_{(n)}, \mathbf{y})$$

at given point $\mathbf{x} \in \mathbf{X}$, the nonconformity measure can be defined as

$$A^*(\mathbf{Z}_{(n)}; (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) \equiv a^*(\mathbf{y}|\mathbf{x}, \mathbf{Z}_{(n)}) = d(\mathbf{f}^*(\mathbf{x}|\mathbf{Z}_{(n)}), \mathbf{y}),$$

where $d(\mathbf{y}^*, \mathbf{y})$ is a chosen proximity measure between estimator $\mathbf{y}^*$ and output $\mathbf{y}$ (for example, $\delta(\mathbf{x}|\mathbf{y})$ from (2)).

Then, the nonconformity measure can be computed for all *old* examples Z_i relative to the examples $\mathbf{Z}_{(n)/i}$, consisting of examples $\mathbf{Z}_{(n)}$ (1) without the i-th example Z_i and *new* example $Z = (\mathbf{x}, \mathbf{y})^{\mathrm{T}}$:

$$A_i \equiv A(\mathbf{Z}_{(n)/i} \cup Z; Z_i) = a_i(\mathbf{y}) = a(\mathbf{y}_i|\mathbf{x}_i, \mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}).$$

After that, given some confidence level, we can compute the frequencies

$$p(\mathbf{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\{a_i(\mathbf{y}) \geq a(\mathbf{y})\},$$

which depend on unknown output $\mathbf{y}$. Note that in Shafer and Vovk (2008) the frequency is computed in a slightly different way, namely as $\left(\frac{1}{n+1} + \frac{1}{n+1} \times p(\mathbf{y})\right)$.

Finally, the value $\mathbf{y}$ is included in the sought-for prediction region $\Gamma_\alpha(\mathbf{x}) = \Gamma_\alpha(\mathbf{x}|\mathbf{Z}_{(n)})$ if and only if $p(\mathbf{y}) > \alpha$.

1.4. Proposed nonconformity measure

In this paper, we introduce a nonconformity measure for estimating the proximity between examples as follows. Let us consider regression problems in which both an unknown mapping $\mathbf{f}(\mathbf{x})$ and its Jacobian matrix $\mathbf{J_f}(\mathbf{x})$ should be estimated.

Let $\mathbf{x}, \mathbf{x}' \in \mathbf{X}$ be some nearby inputs, and $\mathbf{y} = \mathbf{f}(\mathbf{x})$, $\mathbf{y}' = \mathbf{f}(\mathbf{x}')$ are the corresponding outputs. Consider a prediction

$$\hat{\mathbf{y}}' = \mathbf{y} - \mathbf{J_f}(x') \times (\mathbf{x} - \mathbf{x}') \tag{7}$$

of output $\mathbf{y}'$ from given output $\mathbf{y}$ with an error of order $O(|\mathbf{x}' - \mathbf{x}|^2)$.

Let us define a nonconformity measure between examples $Z = (\mathbf{x}, \mathbf{y})^{\mathrm{T}}$ and $Z' = (\mathbf{x}', \mathbf{y}')^{\mathrm{T}}$ in terms of quality of the prediction (7) by asymmetrical measure

$$|\hat{\mathbf{y}}' - \mathbf{y}| = |(\mathbf{y}' - \mathbf{y}) + \mathbf{J_f}(\mathbf{x}) \times (\mathbf{x} - \mathbf{x}')|; \tag{8}$$

this measure is closely related to the Bregman distance Bregman (1967)

$$D_\phi(\mathbf{x}', \mathbf{x}) = \phi(\mathbf{x}) - \phi(\mathbf{x}') - (\nabla\phi(\mathbf{x}'), \mathbf{x}' - \mathbf{x}),$$

defined with the use of continuously-differentiable real-valued and strictly convex function ϕ with closed convex domain of definition (thus, $D_\phi(\mathbf{x}', \mathbf{x}) > 0$), which are widely used in Machine Learning Amari (1998); Amari and Nagaoka (2000); Banerjee et al. (2005a,b).

Taking into account that the measure (8) is defined only for nearby inputs $\mathbf{x}'$ and $\mathbf{x}$, we define value of the nonconformity measure $B(\mathbf{Z}_{(n)}; Z)$ at point $Z = (\mathbf{x}, \mathbf{y})^{\mathrm{T}}$ based on MLR estimators $\hat{\mathbf{y}}(\mathbf{x})$ and $\mathbf{G}_{MLR}(\mathbf{x})$ as follows

$$\begin{aligned} b(\mathbf{x}, \mathbf{y}) = b(\mathbf{x}, \mathbf{y}|\mathbf{Z}_{(n)}) &= \sum_{j=1}^{n} K(\mathbf{x}_j, \mathbf{x}) \times |\hat{\mathbf{y}}(\mathbf{x}_j) - \mathbf{y}_j| = \\ &= \sum_{j=1}^{n} K(\mathbf{x}_j, \mathbf{x}) \times |\mathbf{y}_j - \mathbf{y} + \mathbf{G}_{MLR}(\mathbf{x}_j) \times (\mathbf{x}_j - \mathbf{x})|, \end{aligned} \tag{9}$$

where $\mathbf{G}_{MLR}(\mathbf{x}_j) = \mathbf{G}_{MLR}(\mathbf{x}_j|\mathbf{Z}_{(n)})$ is MLR-estimator for Jacobian matrix $\mathbf{J_f}(\mathbf{x}_j)$ and $K(\mathbf{x}', \mathbf{x}) = \mathrm{I}\{|\mathbf{x}' - \mathbf{x}| < \epsilon\}$ is a kernel determined by a pre-selected small parameter ϵ.

Let us denote

$$\begin{aligned} b_i(\mathbf{x}, \mathbf{y}) &= b_i(\mathbf{x}, \mathbf{y}|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) = \\ &= \sum_{j=1, j\neq i}^{n} K(\mathbf{x}_j, \mathbf{x}_i) \times |\mathbf{y}_j - \mathbf{y}_i + \mathbf{G}_{MLR}(\mathbf{x}_j|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) \times (\mathbf{x}_j - \mathbf{x}_i)| + \\ &+ K(\mathbf{x}, \mathbf{x}_i) \times \left|\mathbf{y} - \mathbf{y}_i + \mathbf{G}_{MLR}(\mathbf{x}|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) \times (\mathbf{x} - \mathbf{x}_i)\right|. \end{aligned} \tag{10}$$

The paper constructs a prediction region $\Gamma_\alpha(\mathbf{x})$ for unknown output $\mathbf{y} = \mathbf{f}(\mathbf{x})$ at OoS input point $\mathbf{x}$ given some confidence level based on the MLR-estimators $\hat{\mathbf{y}}(\mathbf{x})$ and $\mathbf{G}_{MLR}(\mathbf{x})$ using introduced nonconformity measures (9), (10).

The paper is organized as follows. Section 2 describes shortly the MLR algorithm. Section 3 contains an approximation of measures $b(\mathbf{x}, \mathbf{y})$ (9) and $\{b_i(\mathbf{x}, \mathbf{y}), i = 1, 2, \ldots, n\}$ (10), computed for all sample points, and Section 4 describes obtained prediction region.

2. Manifold Learning Regression

2.1. Manifold Estimation Problem

The MLR is based on the estimation of unknown RM $\mathbf{M}(\mathbf{f})$ (3) from given points $\mathbf{Z}_{(n)}$ (1), sampled from the manifold. Considered Manifold estimation problem Kuleshov and Bernstein (2014a,b) for the RM is to construct two interrelated mappings

- an embedding mapping $h : \mathbf{M}(\mathbf{f}) \to \mathbf{U}_h \subset \mathbb{R}^q$ from the RM $\mathbf{M}(\mathbf{f})$ to the Feature space (FS) $\mathbf{U}_h = h(\mathbf{M}(\mathbf{f})) \subset \mathbb{R}^q$, which produces a low-dimensional parameterization (low-dimensional coordinates) $h(Z)$ of manifold points $Z \in \mathbf{M}(\mathbf{f})$,
- a recovery mapping $g : \mathbf{U}_h \to \mathbb{R}^p$ which recovers manifold points $Z \in \mathbf{M}(\mathbf{f})$ from their low-dimensional coordinates $\mathbf{u} = h(Z)$,

which ensure small recovery error (a measure of a quality of solution (h, g) at point Z)

$$\delta_{h,g}(Z) = |r_{h,g}(Z) - Z|, \tag{11}$$

here the recovered value

$$r_{h,g}(Z) = g(h(Z)) \approx Z \tag{12}$$

is a result of successively applying the embedding and recovery mappings to vector Z.

Mappings (h, g) determine q-dimensional recovered Regression manifold (RRM)

$$\begin{aligned}\mathbf{M}_{h,g} &= r_{h,g}(\mathbf{M}(\mathbf{f})) = \{r_{h,g}(Z) \in \mathbb{R}^p : Z \in \mathbf{M}(\mathbf{f})\} \\ &= \{Z = g(\mathbf{u}) \in \mathbb{R}^p : \mathbf{u} \in \mathbf{U}_h = \mathbf{M}(\mathbf{f}) \subset \mathbb{R}^q\},\end{aligned} \tag{13}$$

which is embedded in ambient input space $\mathbb{R}^p$, covered by a single chart g, and consists of all recovered values $r_{h,g}(Z)$ of manifold points Z. Proximities (12) imply manifold proximity $\mathbf{M}_{h,g} \approx \mathbf{M}(\mathbf{f})$ meaning small Hausdorff distance $d_H(\mathbf{M}_{h,g}, \mathbf{M}(\mathbf{f}))$ between the RM $\mathbf{M}(\mathbf{f})$ and RRM $\mathbf{M}_{h,g}$ (13) due to inequality $d_H(\mathbf{M}_{h,g}, \mathbf{M}(\mathbf{f})) \leq \sup_{Z \in \mathbf{M}(\mathbf{f})} \delta_{h,g}(Z)$.

The proximity $r_{h,g}(Z) \approx Z$ can be directly checked at sample point $Z \in \mathbf{Z}_n$; for OoS point $Z \in \mathbf{M}(\mathbf{f}) \setminus \mathbf{Z}_{(n)}$ it describes the generalization ability of solution (h, g) at a specific point Z. To provide a good generalization ability, certain additional requirements to the pair (h, g) have to be satisfied Bernstein and Kuleshov (2013): the pair (h, g) should preserve a differential structure of the RM providing proximity between tangent spaces to the RM $\mathbf{M}(\mathbf{f})$ (3) and RRM $\mathbf{M}_{h,g}$ (13).

In manifold theory Jost (2002); Lee (2009), the set composed of manifold points equipped by tangent spaces at these points is called the Tangent bundle of the manifold. Thus, a problem of manifold recovery, which includes a recovery of its tangent spaces as a subproblem, is referred to as the Tangent bundle manifold learning problem and can be formulated as follows Bernstein and Kuleshov (2012): to construct the solution (h, g) of the manifold estimation problem which, additionally to manifold proximities (11), provides tangent proximity

$$\mathrm{L}_{h,g}(Z) \approx \mathrm{L}(Z), \tag{14}$$

where $\mathrm{L}(Z)$ is a tangent space to the RM $\mathbf{M}(\mathbf{f})$ at the point Z and q-dimensional linear space

$$\mathrm{L}_{h,g}(Z) = \mathrm{Span}(\mathbf{J}_g(h(Z))),$$

spanned by columns of the Jacobian matrix $\mathbf{J}_g(\mathbf{u}), \mathbf{u} = h(Z)$ of the recovery mapping $g(\mathbf{u})$, is a tangent space to the RRM $\mathbf{M}_{h,g}$ (13) at recovered point $r_{h,g}(Z) \in \mathbf{M}_{h,g}$; the proximity in (14) is defined in terms of chosen distance between these tangent spaces, considered as elements of the Grassmann manifold $\mathrm{Grass}(p, q)$, consisting of all q-dimensional linear subspaces in $\mathbb{R}^p$.

2.2. Grassmann & Stiefel Eigenmaps algorithm

Grassmann & Stiefel Eigenmaps (GSE) algorithm Kuleshov and Bernstein (2014b); Bernstein and Kuleshov (2012) provides the solution to the Tangent bundle manifold learning problem and consists of three successively performed steps: tangent manifold learning, manifold embedding, and tangent bundle recovery.

In the first step, a sample-based family $\mathbf{H}$ consisting of $p \times q$ matrices $\mathbf{H}(Z)$ with columns $\{\mathbf{H}^{(k)}(Z) \in \mathbb{R}^p, k = 1, 2, \ldots, q\}$, smoothly depending on Z, is constructed to meet the relations $\mathrm{L}_{\mathbf{H}}(Z) \equiv \mathrm{Span}(\mathbf{H}(Z)) \approx \mathrm{L}(Z)$ and

$$\nabla_{\mathbf{H}^{(i)}(Z)} \mathbf{H}^{(i)}(Z) = \nabla_{\mathbf{H}^{(j)}(Z)} \mathbf{H}^{(i)}(Z), \;\; i, j = 1, 2, \ldots, q, \tag{15}$$

for all $Z \in \mathbf{M}(\mathbf{f})$; covariant differentiation is used in these relations. The mappings h and g will be built on the next steps in such a way that

$$\mathbf{J}_g(h(Z)) = \mathbf{H}(Z), \tag{16}$$

and, therefore, the columns $\{\mathbf{H}^{(k)}(Z)\}$ are coordinate tangent fields on the RM $\mathbf{M}(\mathbf{f})$.

In the next step, given the family $\mathbf{H}$ already constructed, an embedding mapping $h(Z)$ is constructed to meet the relations $Z' - Z \approx \mathbf{H}(Z) \times (h(Z') - h(Z))$ for nearby points $Z', Z \in \mathbf{M}(\mathbf{f})$; the FS $\mathbf{U}_h = h(\mathbf{M}(\mathbf{f}))$ is also determined on this step.

In the final step, given the family $\mathbf{H}$ and already constructed mapping h, a recovery mapping $g(\mathbf{u}), \mathbf{u} \in \mathbf{U}_h$ is constructed to meet the relations (12) and (16).

The estimator $\mathbf{M}_{h,g}$, constructed in this way, meets required manifold and tangent proximities.

2.3. Manifold Learning Regression algorithm

The GSE applied to the sample $\mathbf{Z}_{(n)}$ (1) from the RM $\mathbf{Z}(\mathbf{f})$ results in Embedding mapping $\mathbf{u} = h(Z) \in \mathbb{R}^q, Z \in \mathbb{R}^p$, recovery mapping $Z = g(\mathbf{u}) \in \mathbb{R}^p$, and $p \times q$ matrix $G(\mathbf{u})$, which approximates Jacobian matrix $\mathbf{J}_g(\mathbf{u})$ of the recovery mapping $g(\mathbf{u})$; these quantities fulfill proximity relations (17) and (19).

Thus, we have two parameterizations of manifold points $Z = \mathbf{F}(\mathbf{x}) \in \mathbf{M}(\mathbf{f})$: *natural* parameterization by input $\mathbf{x} \in \mathbf{X}$ and GSE-parameterization $\mathbf{u} = h(Z)$, which are linked by an unknown one to one *reparameterization* mapping

$$\mathbf{u} = \varphi(\mathbf{x}) : \mathbf{x} \in \mathbf{X} \to \varphi(\mathbf{x}) = h(\mathbf{F}(\mathbf{x})) \in \mathbf{U}_h. \tag{17}$$

A splitting of the p-dimensional vector $Z = \begin{pmatrix} Z_{in} \\ Z_{out} \end{pmatrix}$, $p = q + m$, on the q-dimensional vector Z_{in} and m-dimensional vector Z_{out} implies the corresponding partitions

$$g(\mathbf{y}) = \begin{pmatrix} g_{in}(\mathbf{u}) \\ g_{out}(\mathbf{u}) \end{pmatrix} \text{ and } \mathbf{G}(\mathbf{y}) = \begin{pmatrix} \mathbf{G}_{in}(\mathbf{u}) \\ \mathbf{G}_{out}(\mathbf{u}) \end{pmatrix} \tag{18}$$

of the vector $g(\mathbf{y}) \in \mathbb{R}^p$ and $p \times q$ matrix $\mathbf{G}(\mathbf{y})$; the $q \times q$ and $m \times q$ matrices $\mathbf{G}_{in}(\mathbf{u})$ and $\mathbf{G}_{out}(\mathbf{u})$ are the Jacobian matrices of the mappings $g_{in}(\mathbf{u})$ and $g_{out}(\mathbf{u})$. It follows from proximities (12), (14) with the use of representation $Z = \mathbf{F}(\mathbf{X}) = \begin{pmatrix} \mathbf{x} \\ \mathbf{f}(\mathbf{x}) \end{pmatrix}$, that

$$g_{in}(h(\mathbf{F}(\mathbf{x}))) \approx \mathbf{x},$$
$$g_{out}(h(\mathbf{F}(\mathbf{X}))) \approx \mathbf{f}(\mathbf{x}).$$

Therefore,

$$g_{in}(\varphi(\mathbf{x})) \approx \mathbf{x}, \tag{19}$$
$$g_{out}(\varphi(\mathbf{x})) \approx \mathbf{f}(\mathbf{x}), \tag{20}$$

and, if the mapping $\mathbf{u} = \varphi(\mathbf{x})$ (17) would be known, the function $\mathbf{f}^*(\mathbf{x}) = g_{out}(\varphi(\mathbf{x}))$ would be sought-for learned function for unknown function $\mathbf{f}(\mathbf{x})$ with small error $\delta(\mathbf{x})$ (2).

Unknown function $\mathbf{u} = \varphi(\mathbf{x})$ (17) is known at sample input points $\{\mathbf{x}_i\}$:

$$\varphi(\mathbf{x}_i) = h(Z_i) \equiv \mathbf{u}_i,\ i = 1, 2, \ldots, n. \tag{21}$$

The GSE-solution gives also known values of Jacobian matrix $\mathbf{J}_\varphi(\mathbf{x})$ of the mapping $\varphi(\mathbf{x})$ at sample input points $\{\mathbf{x}_i\}$: differentiating approximate identity $g_{in}(\varphi(\mathbf{x})) \approx \mathbf{x}$ (19) and using GSE-estimator $\mathbf{G}_{in}(\mathbf{u})$ for the Jacobian matrix of the mapping $g_{in}(\mathbf{u})$, we obtain

$$\mathbf{G}_{in}(\varphi(\mathbf{x})) \times \mathbf{J}_\varphi(\mathbf{x}) \approx \mathrm{I}_q,$$

whence follows

$$\mathbf{J}_\varphi(\mathbf{x}) \approx \mathbf{G}_\varphi(\mathbf{x}) \equiv \mathbf{G}_{in}^{-1}(\varphi(\mathbf{x})), \tag{22}$$

therefore, under known values $\mathbf{G}_{in}(\varphi(\mathbf{x}_i)) = \mathbf{G}_{in}(h(Z_i)) = G_{in}(\mathbf{u}_i)$,

$$\mathbf{J}_\varphi(\mathbf{x}_i) \approx \mathbf{G}_\varphi(\mathbf{x}_i) \equiv \mathbf{G}_{in}^{-1}(\mathbf{u}_i),\ \ i = 1, \ldots, n. \tag{23}$$

Therefore, we obtain a problem called Regression problems with known Jacobians Kuleshov and Bernstein (2016), consisting in estimating an unknown mapping $\varphi(\mathbf{x}) : \mathbb{R}^q \to \mathbb{R}^q$ from its known values $\{\varphi(\mathbf{x}_i) = \mathbf{u}_i\}$ (21) under known values $\{\mathbf{J}_\varphi(\mathbf{x}_i) \approx \mathbf{G}_\varphi(\mathbf{x}_i)\}$ (23) of its $q \times q$ Jacobian matrix $\mathbf{J}_\varphi(\mathbf{x})$ at sample input points $\{\mathbf{x}_i \in \mathbf{X}_{(n)}\}$. The solution to this problem $U(\mathbf{x})$ called Known Jacobian Regression (KJR), given in Kuleshov and Bernstein (2016), is

$$U(\mathbf{x}) = \frac{1}{K(\mathbf{x})}\sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x}) \times \{\mathbf{u}_i + \mathbf{G}_\varphi(\mathbf{x}_i) \times (\mathbf{x} - \mathbf{x}_i)\}. \tag{24}$$

Here $K(\mathbf{x}) = \sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x})$.

Thus, the MLR-solution to the initial estimation problem is given by formula

$$\mathbf{f}_{MLR}(\mathbf{x}) = g_{out}(U(\mathbf{x})).$$

Differentiating approximate identity $g_{out}(\varphi(\mathbf{x})) \approx \mathbf{f}(\mathbf{x})$ (20) and using GSE-estimator $\mathbf{G}_{out}(\mathbf{u})$ for the Jacobian matrix of the mapping $g_{out}(\mathbf{u})$, we obtain

$$\mathbf{G}_{out}(\varphi(\mathbf{x})) \times \mathbf{J}_\varphi(\mathbf{x}) \approx \mathbf{J}_\mathbf{f}(\mathbf{x}),$$

whence follows, using estimators $U(\mathbf{x})$ (24) and $\mathbf{G}_\varphi(\mathbf{x})$ (22) for $\varphi(\mathbf{x})$ and $\mathbf{J}_\varphi(\mathbf{x})$, that

$$\mathbf{G}_{MLR}(\mathbf{x}) \equiv \mathbf{G}_{out}(U(\mathbf{x})) \times \mathbf{G}_{in}^{-1}(U(\mathbf{x})). \tag{25}$$

is MLR-estimator for the unknown Jacobian matrix $\mathbf{J}_\mathbf{f}(\mathbf{x})$ of the mapping $\mathbf{f}$.

In performed numerical experiments Bernstein et al. (2015); Kuleshov and Bernstein (2016), the proposed MLR method compared favorably with the kernel nonparametric regression estimator with stationary kernel.

3. Approximation of proposed nonconformity measure for Manifold Learning Regression

In accordance with common Conformal prediction technique (see Section 1.3), the computed functions $b(\mathbf{x}, \mathbf{y}) = b(\mathbf{x}, \mathbf{y}|\mathbf{Z}_{(n)})$ (9) and $b_i(\mathbf{x}, \mathbf{y}) = b_i(\mathbf{x}, \mathbf{y}|Z_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}})$ (10) allows constructing a sought-for prediction region $\Gamma(\mathbf{x}) = \Gamma_{\alpha,MLR}(\mathbf{x}|\mathbf{Z}_{(n)})$ as follows: the point $\mathbf{y}$ is included in the region if and only if $\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\{b_i(\mathbf{x}, \mathbf{y}) \geq b(\mathbf{x}, \mathbf{y})\} > \alpha$.

Taking into account proximity $\mathbf{f}_{MLR}(\mathbf{x}) - \mathbf{f}(\mathbf{x}) = O(n^{-2/(q+2)})$ (4), we will consider values $\mathbf{y}$ which is located "not very far" asymptotically from the MLR-estimator $\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{f}_{MLR}(\mathbf{x})$ and belong to the ball

$$W_{C,\gamma}(\hat{\mathbf{y}}(\mathbf{x})) = \{\mathbf{y} : |\mathbf{y} - \hat{\mathbf{y}}(\mathbf{x})| < C \times n^{-(1+\gamma)/(q+2)}\} \tag{26}$$

centered at $\hat{\mathbf{y}}(\mathbf{x})$ with radius $C \times n^{-(1+\gamma)/(q+2)}$, where C and γ, $0 < \gamma \leq 1$, are chosen parameters.

The quantities (9) and (10) depend on MLR-estimators $\mathbf{G}_{MLR}(\mathbf{x}'|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}})$, constructed from dataset at the points $\mathbf{x}' = \mathbf{x}_j$, $j = 1, 2, \ldots, n$, $j \neq i$, and $\mathbf{x}' = \mathbf{x}$.

As was said above (Section 2.2), MLR-estimator $\mathbf{G}_{MLR}(\mathbf{x})$ (25) for the unknown Jacobian matrix $\mathbf{J}_\mathbf{f}(\mathbf{x})$ of the mapping $\mathbf{f}$ at chosen point $\mathbf{x}$ is based on approximation $\mathbf{G}(\mathbf{u})$ of $p \times q$ Jacobian matrix $\mathbf{J}_g(\mathbf{u})$ of recovery mapping $g(\mathbf{u})$ which is a result of applying the GSE algorithm to the sample $\mathbf{Z}_{(n)}$ (1). Matrix $\mathbf{G}(\mathbf{u})$, in turn, is based on constructing of $p \times q$ the matrices $\mathbf{H}(Z)$ with required properties in the GSE tangent manifold learning step; the mappings $h(Z)$ and $g(\mathbf{u})$ are constructed in the next steps in such a way as that

$\mathbf{J}_g(h(Z)) = \mathbf{H}(Z)$. Note that the values Z are known for sample points; at OoS point, when output $\mathbf{y} = \mathbf{f}(\mathbf{x})$ is unknown, certain interpolation procedure is used in the GSE.

The MLR-estimators $\mathbf{G}_{MLR}(\mathbf{x})$ should be computed at various points $\mathbf{x}' \in \mathbf{X}$ from the sample $\mathbf{Z}_{(n)}$, as well as from modified sample $\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}$. To avoid using a "brute force" in computing these estimators, asymptotic approximations for these estimators, in which a dependency on $\mathbf{y}$ is written in explicit form, is constructed in next subsections for big sample size n. This allows constructing the prediction region "almost" in explicit form.

We start with constructing the approximation of the estimator $\mathbf{G}_{MLR}(\mathbf{x}'|\mathbf{Z}_{(n)})$ from the initial sample $\mathbf{Z}_{(n)}$.

3.1. Approximation for local covariance matrix based on initial sample

The matrices $\mathbf{H}(Z)$ are constructed in the GSE as follows. Let $Z' = \mathbf{F}(\mathbf{x}) \in \mathbf{Z}_{(n)}$, see (3), be chosen reference point and dataset

$$S_\varepsilon(\mathbf{x}) = S_\varepsilon(\mathbf{x}'|\mathbf{Z}_{(n)}) = \{Z_j \in \mathbf{Z}_{(n)} : |\mathbf{x}_j - \mathbf{x}'| < \varepsilon\} \tag{27}$$

consists of sample points whose inputs belong to ε-ball centered at $\mathbf{x}'$ for chosen small parameter ε.

The tangent space $\mathrm{L}(Z)$ at the point $Z' \in \mathbf{M}(\mathbf{f})$ is estimated by the q-dimensional linear space $\mathrm{L}_{PCA}(Z')$ spanned by the eigenvectors of local sample covariance matrix

$$\Sigma_n(Z'|\mathbf{Z}_{(n)}) = \frac{1}{n(\mathbf{x}')} \sum_{j=1}^{n} K(\mathbf{x}', \mathbf{x}_j) \times \left[(Z_j - Z') \times (Z_j - Z')^{\mathrm{T}}\right] \tag{28}$$

corresponding to the q largest eigenvalues which are a result of the Principal Component Analysis (PCA), applied to the points from the dataset $S_\varepsilon(Z)$, where

$$n(\mathbf{x}') = \sum_{j=1}^{n} K(\mathbf{x}', \mathbf{x}_j) \tag{29}$$

is a random number of sample point $\mathbf{x}_j \in \mathbf{X}_{(n)}$ fallen in ε-ball of considered point $\mathbf{x}$.

In this section we study an asymptotic approximation of the local sample covariance matrix $\Sigma_n(Z'|\mathbf{Z}_{(n)})$ (28).

The PCA principal vectors form $p \times q$ matrix $Q_{PCA}(Z') = Q_{PCA}(Z'|\mathbf{Z}_{(n)})$, whose columns are these principal eigenvectors which, in turn, form orthogonal basis in the linear spaces $\mathrm{L}_{PCA}(Z')$. Therefore,

$$\Sigma_n(Z'|\mathbf{Z}_{(n)}) = Q_{PCA}(Z') \times \Lambda_n(Z') \times Q_{PCA}^{\mathrm{T}}(Z'), \tag{30}$$

where

$$\Lambda_n(Z') = \begin{pmatrix} \Lambda_{n,q}(Z') & \mathbf{0}_{qm} \\ \mathbf{0}_{mq} & \Lambda_{n,m}(Z') \end{pmatrix}$$

is $(q+m) \times (q+m)$ diagonal matrix in which diagonal elements of $q \times q$ submatrix $\Lambda_{n,q}(Z')$ consists of q largest eigenvalues and diagonal elements of $m \times m$ submatrix $\Lambda_{n,m}(Z)$ consists of remaining eigenvalues; hereinafter $\mathbf{0}_{st}$ denotes null $s \times t$ matrix.

Assume that the input points (6) are randomly and independently of each other sampled from the Input space $\mathbf{X}$ according to an unknown measure in $\mathbb{R}^q$ with unknown density $\mu(\mathbf{x})$ whose support coincides with the $\mathbf{X}$.

If the RM $\mathbf{M}(\mathbf{f})$ is "well sampled" (sample size n is large enough) and ε is small enough (in subsequent considerations, ε depend on n such as $\varepsilon = O(n^{-1/(q+2)})$) then Levina and Bickel (2005)

$$n(\mathbf{x}') = n \times \mu(\mathbf{x}') \times V_q \times \varepsilon^q \times (1 + O(\varepsilon)) = O(n^{2/(q+2)}), \tag{31}$$

here $V_q = \frac{\pi^{q/2}}{\Gamma(q/2+1)}$ is the volume of q-dimensional unit ball in $\mathbb{R}^q$. Hereinafter, symbol $O(\cdot)$, depending on its use in formulas, refers to a number, vector or matrix, respectively, and $O(\varepsilon^k)$ denotes $O(n^{-k/(q+2)})$. Note that $\mathrm{L}_{PCA}(Z) \approx \mathrm{L}(Z)$ with an error $O(\varepsilon)$ Singer and Wu; Tyagi et al. (2012); Kaslovsky and Meyer (2011); Yanovich (2016).

Under condition $|\mathbf{x}' - \mathbf{x}_j| < \varepsilon$ and choice $\varepsilon = O(n^{-1/(q+2)})$, Taylor series expansion yields

$$\mathbf{F}(\mathbf{x}_j) - \mathbf{F}(\mathbf{x}') = \mathbf{J}_\mathbf{F}(\mathbf{x}') \times (\mathbf{x}_j - \mathbf{x}') + O(\varepsilon^2) = \mathbf{J}_\mathbf{F}(\mathbf{x}') \times (\mathbf{x}_j - \mathbf{x}') + O(n^{-2/(q+2)}),$$

where $p \times q$ matrix $\mathbf{J}_\mathbf{F}(\mathbf{x})$ is

$$\mathbf{J}_\mathbf{F}(\mathbf{x}') = \begin{pmatrix} \mathrm{I}_q \\ \mathbf{J}_\mathbf{f}(\mathbf{x}') \end{pmatrix},$$

therefore,

$$\begin{aligned}\Sigma_n(Z'|\mathbf{Z}_{(n)}) &= \frac{1}{n(\mathbf{x}')} \sum_{j=1}^{n} K(\mathbf{x}', \mathbf{x}_j) \times (\mathbf{F}(\mathbf{x}_j) - \mathbf{F}(\mathbf{x}')) \times (\mathbf{F}(\mathbf{x}_j) - \mathbf{F}(\mathbf{x}'))^{\mathrm{T}} \\ &= \Sigma_{\mathbf{x},n}(\mathbf{x}'|\mathbf{Z}_{(n)}) + O(\varepsilon^3),\end{aligned}$$

where

$$\begin{aligned}\Sigma_{\mathbf{x},n}(\mathbf{x}'|\mathbf{X}_{(n)}) &= \frac{1}{n(\mathbf{x}')} \sum_{j=1}^{n} K(\mathbf{x}', \mathbf{x}_j) \times \left[\mathbf{J}_\mathbf{f}(\mathbf{x}') \times (\mathbf{x}_j - \mathbf{x}')\right] \times \left[\mathbf{J}_\mathbf{f}(\mathbf{x}') \times (\mathbf{x}_j - \mathbf{x}')\right]^{\mathrm{T}} \\ &= \mathbf{J}_F(\mathbf{x}') \times \Sigma_{\mathbf{xx},n}(\mathbf{x}'|\mathbf{X}_{(n)}) \times \mathbf{J}_F^{\mathrm{T}}(\mathbf{x}')\end{aligned} \tag{32}$$

and

$$\Sigma_{\mathbf{xx},n} = \frac{1}{n(\mathbf{x})} \sum_{i=1}^{n} K(\mathbf{x}, \mathbf{x}_i) \times \left[(\mathbf{x}_i - \mathbf{x}) \times (\mathbf{x}_i - \mathbf{x})^{\mathrm{T}}\right]. \tag{33}$$

Matrix $\Sigma_{\mathbf{x},n}(\mathbf{x}'|\mathbf{X}_{(n)})$ (32) has rank q, hence it has no more than q nonzero eigenvalues, and relation (30) can be written as

$$\Sigma_n(Z'|\mathbf{Z}_{(n)}) = Q_{PCA}(Z') \times \Lambda_{n,q}(Z') \times Q_{PCA}^{\mathrm{T}}(Z') + O(\varepsilon^3). \tag{34}$$

As was shown in (40), under an assumption about a random nature of the sample, random sample point $\mathbf{x}_j \in \mathbf{X}_n$ which is fallen into the set $S_\varepsilon(\mathbf{x})$ (27) has conditional asymptotically ($n \to \infty$ and $\varepsilon \to 0$) uniform distribution in the ε-ball centered at $\mathbf{x}'$.

Therefore, $q \times q$ random matrix $\varepsilon^{-2} \times \Sigma_{\mathbf{xx},n}(\mathbf{x}', \mathbf{X}_{(n)})$ (33) having random number $n(\mathbf{x}')$ (29) summands, by the law of large numbers converge to the $q \times q$ covariance matrix $\mathbb{E}(\xi \times \xi^{\mathrm{T}})$ of q-dimensional random vector ξ uniformly distributed in the unit ball in $\mathbb{R}^q$ with convergence rate $O(n^{-1/2}(\mathbf{x})) = O(n^{-1/(q+2)}) = O(\varepsilon)$. Diagonal matrix $\mathbb{E}(\xi \times \xi^{\mathrm{T}})$ has identical diagonal elements $\frac{q}{q+2}$ which is a mean value of squared length of q-dimensional random vector uniformly distributed in the unit ball in $\mathbb{R}^q$. Therefore,

$$\Sigma_{\mathbf{xx},n}(\mathbf{x}'|\mathbf{X}_{(n)}) = \varepsilon^2 \times \sigma_q^2 \times \mathrm{I}_q \times O(\varepsilon^3),$$

here $\sigma_q^2 = \mu \times \frac{q}{q+2}$, whence, due to relations (32),

$$\begin{aligned}\Sigma_n(Z'|\mathbf{Z}_{(n)}) &= \left[\mathbf{J}_{\mathbf{F}}(\mathbf{x}') \times \Sigma_{\mathbf{xx},n}(\mathbf{x}'|\mathbf{X}_{(n)}) \times \mathbf{J}_{\mathbf{F}}^{\mathrm{T}}(\mathbf{x}')\right] + O(\varepsilon^3) \\ &= \varepsilon^2 \times \sigma_q^2 \times \left[\mathbf{J}_{\mathbf{F}}(\mathbf{x}') \times \mathbf{J}_{\mathbf{F}}^{\mathrm{T}}(\mathbf{x}')\right] + O(\varepsilon^3). \end{aligned} \tag{35}$$

Taking into account the relations (4) and (35), we obtained other asymptotic representations for sample covariance matrix $\Sigma_{\mathbf{xx},n}(\mathbf{x}, \mathbf{X}_{(n)})$ (28):

$$\Sigma_n(Z'|\mathbf{Z}_{(n)}) = \varepsilon^2 \times \sigma_q^2 \times \mathbf{G}_F(\mathbf{x}') + O(\varepsilon^3),$$

where

$$\mathbf{G}_{\mathbf{F}}(\mathbf{x}') = \begin{pmatrix} \mathrm{I}_q & \mathbf{G}_{MLR}^{\mathrm{T}}(\mathbf{x}) \\ \mathbf{G}_{MLR}(\mathbf{x}) & \mathbf{G}_{MLR}(\mathbf{x}) \times \mathbf{G}_{MLR}^{\mathrm{T}}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \mathrm{I}_q \\ \mathbf{G}_{MLR}(\mathbf{x}) \end{pmatrix} \times \begin{pmatrix} \mathrm{I}_q \\ \mathbf{G}_{MLR}(\mathbf{x}) \end{pmatrix}^{\mathrm{T}}. \tag{36}$$

3.2. Details of constructing the MLR-estimators for Jacobian matrices

The PCA principal vectors as orthogonal bases in the linear spaces $\mathrm{L}_{PCA}(Z')$ are not agreed by with each other and can be very different, even in close points. GSE constructs other bases in these spaces, which form $p \times q$ matrices

$$\mathbf{H}(Z') = Q_{PCA}(Z') \times v(Z'), \tag{37}$$

where $q \times q$ nonsingular matrices $v(Z')$ should provide the smooth dependency $\mathbf{H}(Z')$ at point Z' and equalities (15). As was said above, $p \times q$ matrix $\mathbf{H}(Z')$ equals to Jacobian matrix $\mathbf{G}(\mathbf{u}')$ of the recovery mapping $g(\mathbf{u})$ at the point $\mathbf{u}' = h(Z')$.

In accordance with the partition (18), we split $p \times q$ matrix $Q_{PCA}(Z)$ on $q \times q$ matrix $Q_{PCA,in}(Z')$ and $m \times q$ matrix $Q_{PCA,out}(Z')$:

$$Q_{PCA}(Z') = \begin{pmatrix} Q_{PCA,in}(Z') \\ Q_{PCA,out}(Z') \end{pmatrix},$$

and denote $m \times q$ matrix

$$\mathbf{G}_{PCA}(\mathbf{x}') = Q_{PCA,out}(\mathbf{F}(\mathbf{x}')) \times Q_{PCA,in}^{-1}(\mathbf{F}(\mathbf{x}')). \tag{38}$$

It follows from (16) and (37) that

$$\begin{aligned}\mathbf{G}_{in}(h(Z')) &= Q_{PCA,in}(Z') \times v(Z'), \\ \mathbf{G}_{out}(h(Z')) &= Q_{PCA,out}(Z') \times v(Z').\end{aligned}$$

The MLR-estimator $\mathbf{G}_{MLR}(\mathbf{x}')$ (25), taking into account the relations (16), (37) and that $U(\mathbf{x}') = h(\mathbf{F}(\mathbf{x}'))$ for point $Z' = \mathbf{F}(\mathbf{x}')$, is as follows

$$\begin{aligned}\mathbf{G}_{MLR}(\mathbf{x}') &= \mathbf{G}_{out}(U(\mathbf{x})) \times \mathbf{G}_{in}^{-1}(U(\mathbf{x}'))\\ &= [Q_{PCA,out}(Z') \times v(Z')] \times [Q_{PCA,in}(Z') \times v(Z')]^{-1}\\ &= [Q_{PCA,out}(Z') \times v(Z')] \times [v^{-1}(Z') \times Q_{PCA,in}^{-1}(Z')]\\ &= Q_{PCA,out}(Z') \times Q_{PCA,in}^{-1}(Z').\end{aligned}$$

Therefore, the estimator (25) doesn't depend on matrix $v(Z')$, and:

$$\mathbf{G}_{MLR}(\mathbf{x}') = \mathbf{G}_{MLR}(\mathbf{x}'|\mathbf{Z}_{(n)}) = Q_{PCA,out}(\mathbf{F}(\mathbf{x}')) \times Q_{PCA,in}^{-1}(\mathbf{F}(\mathbf{x}')) = \mathbf{G}_{PCA}(\mathbf{x}'). \quad (39)$$

Estimator $\mathbf{G}_{MLR}(\mathbf{x}|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}})$ of the Jacobian matrix $\mathbf{J}_{\mathbf{f}}(\mathbf{x}')$ computed at point $\mathbf{x}'$, constructed from the data $\{\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}\}$, is equal to

$$\mathbf{G}_{MLR}(\mathbf{x}'|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) = Q_{PCA,out}(Z'|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) \times Q_{PCA,in}^{-1}(Z'|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}), \quad (40)$$

here $(q + m) \times q$ matrix

$$Q_{PCA,i}(\mathbf{x}, \mathbf{y}) = Q_{PCA}\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \mid \mathbf{Z}_{(n),i} \cup \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right) \quad (41)$$

consists of q $(q + m)$-dimensional columns which are the eigenvectors of matrix

$$\begin{aligned}\Sigma_n(Z'|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) = &\frac{1}{n(\mathbf{x}', \mathbf{x}_i, \mathbf{x})} \sum_{j=1, j\neq i}^{n} K(\mathbf{x}', \mathbf{x}_j) \times \left[(Z_j - Z') \times (Z_j - Z')^{\mathrm{T}}\right]\\ &+ \frac{1}{n(\mathbf{x}', \mathbf{x}_i, \mathbf{x})} K(\mathbf{x}', \mathbf{x}) \times \begin{pmatrix}\mathbf{x} - \mathbf{x}'\\ \mathbf{y} - \mathbf{y}'\end{pmatrix} \times \begin{pmatrix}\mathbf{x} - \mathbf{x}'\\ \mathbf{y} - \mathbf{y}'\end{pmatrix}^{\mathrm{T}}, \quad (42)\end{aligned}$$

corresponding to the q largest eigenvalues, and

$$n(\mathbf{x}, \mathbf{x}_i, \mathbf{x}) = n(\mathbf{x}') - K(\mathbf{x}_i, \mathbf{x}') + \mathrm{I}\{|\mathbf{x} - \mathbf{x}'| < \varepsilon\};$$

we will call the matrix (42) modified local sample covariance matrix.

It follows from (31) that

$$n(\mathbf{x}', \mathbf{x}_i, \mathbf{x}) = n(\mathbf{x}') \times (1 + O(\varepsilon^2)) = O(\varepsilon^3). \quad (43)$$

Therefore, we need to compute the estimators $\mathbf{G}_{MLR}(\mathbf{x}|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}})$ (40) at all points $\mathbf{x} = \mathbf{x}_k$, $k \neq i$, and $\mathbf{x}' = \mathbf{x}$ for $i = 1, 2, \ldots, n$. For this, using an identity

$$\begin{pmatrix}\mathbf{x} - \mathbf{x}'\\ \mathbf{y} - \mathbf{y}'\end{pmatrix} = \begin{pmatrix}\mathbf{x} - \mathbf{x}'\\ \mathbf{f}(\mathbf{x}) - \mathbf{y}'\end{pmatrix} + \begin{pmatrix}\mathbf{0}_{1q}\\ \mathbf{f}_{MLR}(\mathbf{x}) - \mathbf{f}(\mathbf{x})\end{pmatrix} + \begin{pmatrix}\mathbf{0}_{1q}\\ \mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})\end{pmatrix} \quad (44)$$

and taking into account property (4), the relation

$$\begin{pmatrix}\mathbf{x} - \mathbf{x}'\\ \mathbf{y} - \mathbf{y}'\end{pmatrix} = \begin{pmatrix}\mathbf{x} - \mathbf{x}'\\ \mathbf{f}(\mathbf{x}) - \mathbf{y}'\end{pmatrix} + \begin{pmatrix}\mathbf{0}_{1q}\\ \mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})\end{pmatrix} + O(\varepsilon^2) \quad (45)$$

will be used for these purposes.

3.3. Approximation for modified local covariance matrix at sample points

In this section we study an asymptotic approximation of modified local sample covariance matrix $\Sigma_n(Z'|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}})$ (42) at sample points $Z' = Z_k$, $k = 1, 2, \ldots, n$, $k \neq i$.

Taking into account approximate relation (45), we obtain

$$\Sigma_n(Z_k|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) = \frac{1}{n(\mathbf{x}_k, \mathbf{x}_i, \mathbf{x})}\Big\{\sum_{j=1, j\neq i}^{n} K(\mathbf{x}_k, \mathbf{x}_j) \times \left[(Z_j - Z_k) \times (Z_j - Z_k)^{\mathrm{T}}\right] +$$
$$+ K(\mathbf{x}_k, \mathbf{x}) \times \left[(\mathbf{F}(\mathbf{x}) - Z_k)(\mathbf{F}(\mathbf{x}) - Z_k)^{\mathrm{T}}\right]\Big\} +$$
$$+ \frac{1}{n(\mathbf{x}_k, \mathbf{x}_i, \mathbf{x})} K(\mathbf{x}_k, \mathbf{x}) \times \begin{pmatrix} \mathbf{0}_{qq} & \Delta_{MLR,\mathbf{x}}(\mathbf{x}, \mathbf{x}_k) \\ \Delta^{\mathrm{T}}_{MLR,\mathbf{x}}(\mathbf{x}, \mathbf{x}_k) & \Delta_{MLR,\mathbf{y}}(\mathbf{x}, \mathbf{x}_k) \end{pmatrix} + O(\varepsilon^2), \tag{46}$$

where

$$\Delta_{MLR,\mathbf{x}}(\mathbf{x}, \mathbf{x}_k) = (\mathbf{x} - \mathbf{x}_k) \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}))^{\mathrm{T}}, \tag{47}$$
$$\Delta_{MLR,\mathbf{y}}(\mathbf{x}, \mathbf{x}_k) = (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})) \times (\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_k))^{\mathrm{T}} + (\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_k)) \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}))^{\mathrm{T}}$$
$$+ (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})) \times (\mathbf{y} - \mathbf{f}_{MLR}(x))^{\mathrm{T}}. \tag{48}$$

The point $\mathbf{F}(\mathbf{x})$ can be considered as an additional sample point, therefore, a quantity in brackets in the first term in (46) coincides with the sum in the local covariance matrix $\Sigma_n(Z_k|\mathbf{Z}_{(n)})$ (28), in which i-th sample point Z_i is replaced by the point $\mathbf{F}(\mathbf{x})$. Thus, taking into account relation (43), the first term in (46) coincides asymptotically with the $\Sigma_n(Z_k|\mathbf{Z}_{(n)})$ (28) up to term $O(\varepsilon^3)$.

Matrices (47), (48) have the order $O(\varepsilon^{2+\gamma})$, hence, second term in (46) has the order $O(\varepsilon^{4+\gamma})$, whence

$$\Sigma_n(Z_k|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) = \Sigma_n(Z_k|\mathbf{Z}_{(n)}) + O(\varepsilon^3). \tag{49}$$

Therefore, using notation (38) and relation (39),

$$\mathbf{G}_{MLR}(\mathbf{x}_k|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) = \mathbf{G}_{PCA}(\mathbf{x}_k) + O(\varepsilon^3). \tag{50}$$

3.4. Approximation for modified local covariance matrix at reference points

A case $Z' = \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}$ is considered in this section. Write the relations (44), (45) as

$$Z_j - \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_j - \mathbf{x} \\ Z_j - \mathbf{f}(\mathbf{x}) \end{pmatrix} - \begin{pmatrix} \mathbf{0}_{1q} \\ \mathbf{f}_{MLR}(\mathbf{x}) - \mathbf{f}(\mathbf{x}) \end{pmatrix} - \begin{pmatrix} \mathbf{0}_{1q} \\ \mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}) \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{x}_j - \mathbf{x} \\ Z_j - \mathbf{f}(\mathbf{x}) \end{pmatrix} - \begin{pmatrix} \mathbf{0}_{1q} \\ \mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}) \end{pmatrix} + O(\varepsilon^2),$$

thus, matrix $\Sigma_n\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} |\mathbf{Z}_{(n)/i} \cup \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right)$ (41) is equal to

$$\Sigma_n\left(\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} |\mathbf{Z}_{(n)/i} \cup \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right) = \frac{1}{n(\mathbf{x}, \mathbf{x}_i)} \sum_{j=1, j\neq i}^{n} K(\mathbf{x}, \mathbf{x}_j) \times \left[(Z_j - \mathbf{F}(\mathbf{x})) \times (Z_j - \mathbf{F}(\mathbf{x}))^{\mathrm{T}}\right]$$
$$+ \begin{pmatrix} \mathbf{0}_{qq} & \Delta_{MLR,\mathbf{x},n,i} \\ \Delta^{\mathrm{T}}_{MLR,\mathbf{x},n,i} & \Delta_{MLR,\mathbf{y},n,i} \end{pmatrix} + O(\varepsilon^3), \tag{51}$$

where

$$n(\mathbf{x}, \mathbf{x}_i) = n(\mathbf{x}) - K(\mathbf{x}_i, \mathbf{x}) = O(\varepsilon^2) \tag{52}$$

and

$$\begin{aligned}
\Delta_{MLR,\mathbf{x},n,i} &= -\left(\overline{\mathbf{x}}_{/i} - \mathbf{x}\right) \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}))^{\mathrm{T}}, \\
\Delta_{MLR,\mathbf{y},n,i} &= -\left(\overline{\mathbf{f}(\mathbf{x})}_{/i} - \mathbf{f}(\mathbf{x})\right) \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}))^{\mathrm{T}} - \left(\overline{\mathbf{f}(\mathbf{x})}_{/i} - \mathbf{f}(\mathbf{x})\right)^{\mathrm{T}} \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})) \\
&\quad + (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})) \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}))^{\mathrm{T}},
\end{aligned} \tag{53}$$

here

$$\left(\overline{\mathbf{x}}_{/i} - \mathbf{x}\right) = \frac{1}{n(\mathbf{x}, \mathbf{x}_i)} \sum_{j=1, j\neq i}^{n} K(\mathbf{x}, \mathbf{x}_j) \times (\mathbf{x}_j - \mathbf{x}), \tag{54}$$

$$\overline{\mathbf{f}(\mathbf{x})}_{/i} - \mathbf{f}(\mathbf{x}) = \frac{1}{n(\mathbf{x}, \mathbf{x}_i)} \sum_{j=1, j\neq i}^{n} K(\mathbf{x}, \mathbf{x}_j) \times (\mathbf{f}(\mathbf{x}_j) - \mathbf{f}(\mathbf{x})). \tag{55}$$

The first term in (51) coincides with the local covariance matrix $\Sigma_{n-1}(\mathbf{F}(\mathbf{x})|\mathbf{Z}_{(n)/i})$ at the point $Z = \mathbf{F}(\mathbf{x})$, constructed from the sample $\mathbf{Z}_{(n)}$ without sample point Z_i. Thus, taking into account relation (52), the first term in (46) coincides asymptotically with the $\Sigma_n(Z|\mathbf{Z}_{(n)})$ (28) up to term $O(\varepsilon^3)$.

Random vectors (54) and (55), which have random number $n(\mathbf{x}, \mathbf{x}_i)$ (52) of summands, by the law of large numbers converge to zero vectors $\mathbf{0}_q$ with convergence error $O(\varepsilon)$. Therefore, second term in (51) has the order $O(\varepsilon^{2(1+\gamma)})$.

Taking into account the relations (4), (25) and

$$\begin{aligned}
\overline{\mathbf{x}}_{/i} - \mathbf{x} &= \overline{\mathbf{x}} - \mathbf{x} + O(\varepsilon^3), \\
\overline{\mathbf{f}(\mathbf{x})}_{/i} - \mathbf{f}(\mathbf{x}) &= \mathbf{J}_{\mathbf{f}}(\mathbf{x}) \times (\overline{\mathbf{x}} - \mathbf{x}) + O(\varepsilon^3),
\end{aligned}$$

where $\overline{\mathbf{x}}$ is sample mean constructed from "whole" sample $\mathbf{X}_{(n)}$ (1), we obtain

$$\begin{aligned}
&\Sigma_n\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix} \Big| \mathbf{Z}_{(n)/i} \cup \begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right) = \Sigma_n(\mathbf{F}(\mathbf{x})|\mathbf{Z}_{(n)}) \\
&- \begin{pmatrix} \mathbf{0}_{qq} & \delta_{MLR}^{\mathrm{T}}(\mathbf{x}) \\ \delta_{MLR}(\mathbf{x}) & \mathbf{G}_{MLR}(\mathbf{x}) \times \delta_{MLR}^{\mathrm{T}}(\mathbf{x}) + \delta_{MLR}(\mathbf{x}) \times \mathbf{G}_{MLR}^{\mathrm{T}}(\mathbf{x}) - \Delta_{MLR} \end{pmatrix} + O(\varepsilon^{2+2\gamma}),
\end{aligned}$$

here

$$\begin{aligned}
\delta_{MLR}(\mathbf{x}) &= (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})) \times (\overline{\mathbf{x}} - \mathbf{x})^{\mathrm{T}}, \\
\Delta_{MLR} &= (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x})) \times (\mathbf{y} - \mathbf{f}_{MLR}(\mathbf{x}))^{\mathrm{T}}
\end{aligned}$$

are $m \times q$ and $m \times q$ matrices, which have the orders $O(\varepsilon^{2+\gamma})$ and $O(\varepsilon^{2+2\gamma})$ respectively.

Taking into account representation (36) and relation

$$\delta_{MLR}(\mathbf{x}) \times \delta_{MLR}^{\mathrm{T}}(\mathbf{x}) = O(\varepsilon^{4+2\gamma}),$$

we obtain

$$\Sigma_n\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}|\mathbf{Z}_{(n)/i}\cup\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right)=\varepsilon^2\times\sigma_q^2(\mathbf{x})\times\begin{pmatrix}\mathrm{I}_q & \mathbf{G}^{\mathrm{T}}_{MLR}(\mathbf{x})\\ \mathbf{G}_{MLR}(\mathbf{x}) & \mathbf{G}_{MLR}(\mathbf{x})\times\mathbf{G}^{\mathrm{T}}_{MLR}(\mathbf{x})\end{pmatrix}$$
$$-\begin{pmatrix}\mathbf{0}_{qq} & \delta^{\mathrm{T}}_{MLR}(\mathbf{x})\\ \delta_{MLR}(\mathbf{x}) & -\mathbf{G}_{MLR}(\mathbf{x})\times\delta^{\mathrm{T}}_{MLR}(\mathbf{x})-\delta_{MLR}(\mathbf{x})\times\mathbf{G}^{\mathrm{T}}_{MLR}(\mathbf{x})+\delta_{MLR}(\mathbf{x})\times\delta^{\mathrm{T}}_{MLR}(\mathbf{x})\end{pmatrix}$$
$$=\varepsilon^2\times\sigma_q^2(\mathbf{x})\times\begin{pmatrix}\mathrm{I}_q\\ \mathbf{G}_{MLR}(\mathbf{x})-\tau(\mathbf{x},\mathbf{y})\end{pmatrix}\times\begin{pmatrix}\mathrm{I}_q\\ \mathbf{G}_{MLR}(\mathbf{x})-\tau(\mathbf{x},\mathbf{y})\end{pmatrix}^{\mathrm{T}}+O(\varepsilon^{2(1+\gamma)}),\tag{56}$$

where $m\times q$ matrix

$$\tau(\mathbf{x},\mathbf{y})=\varepsilon^{-2}\times\sigma_q^{-2}(\mathbf{x})\times(\mathbf{y}-\mathbf{f}_{MLR}(\mathbf{x}))(\overline{\mathbf{x}}-\mathbf{x})^{\mathrm{T}}\tag{57}$$

has the order $O(\varepsilon^{2\gamma})$.

Representation (34) applied to the matrix $\Sigma_n\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}|\mathbf{Z}_{(n)/i}\cup\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right)$ (42), gives relation

$$\Sigma_n\left(\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}|\mathbf{Z}_{(n)/i}\cup\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right)=Q_{PCA,i}(\mathbf{x},\mathbf{y})\times\Lambda_q(\mathbf{x},\mathbf{y})\times Q_{PCA}(\mathbf{x},\mathbf{y})+O(\varepsilon^3),\tag{58}$$

in which $(q+m)\times q$ matrix $Q_{PCA,i}(\mathbf{x},\mathbf{y})$ is defined in (41) and Λ_q is diagonal $q\times q$ matrix consisting of q largest eigenvalues of the matrix (42).

Comparing the representations (26) and (52), we obtain that $(q+m)\times q$ matrix $Q_{PCA,i}(\mathbf{x},\mathbf{y})$ (39) satisfy to relation

$$Q_{PCA,i}(\mathbf{x},\mathbf{y})=\begin{pmatrix}\mathrm{I}_q\\ \mathbf{G}_{MLR}(\mathbf{x})-\tau(\mathbf{x},\mathbf{y})\end{pmatrix}\times S+O(\varepsilon^{2(1+\gamma)}),\tag{59}$$

where

$$S=\varepsilon^2\times\sigma_q^2(\mathbf{x})\times\begin{pmatrix}\mathrm{I}_q\\ \mathbf{G}_{MLR}(\mathbf{x})-\tau(\mathbf{x},\mathbf{y})\end{pmatrix}^{\mathrm{T}}\times Q_{PCA,i}(\mathbf{x},\mathbf{y})\times\Lambda_q^{-1}(\mathbf{x},\mathbf{y})+O(\varepsilon^{2(1+\gamma)}),\tag{60}$$

is $q\times q$ matrix.

The relation (40) and (59), (60) give

$$\mathbf{G}_{MLR}\left(\mathbf{x}\middle|\mathbf{Z}_{(n)/i}\cup\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right)=Q_{PCA,i,out}(\mathbf{x},\mathbf{y})\times Q^{-1}_{PCA,i,in}(\mathbf{x},\mathbf{y}),$$

where

$$Q_{PCA,i,in}(\mathbf{x},\mathbf{y})=S+O(\varepsilon^{2(1+\gamma)}),$$
$$Q_{PCA,i,out}(\mathbf{x},\mathbf{y})=(\mathbf{G}_{MLR}(\mathbf{x})-\tau(\mathbf{x},\mathbf{y}))\times S+O(\varepsilon^{2(1+\gamma)}),$$

whence, with using notation (57),

$$\mathbf{G}_{MLR}\left(\mathbf{x}\middle|\mathbf{Z}_{(n)/i}\cup\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\right)=\mathbf{G}_{MLR}(\mathbf{x})-\tau(\mathbf{x},\mathbf{y})+O(\varepsilon^{1+2\gamma}).\tag{61}$$

PCA-matrix $Q_{PCA}(Z') = Q_{PCA}(Z'|\mathbf{Z}_{(n)})$, and, therefore, the quantity $\mathbf{G}_{PCA}(\mathbf{x}')$ (38) are defined and can be computed in sample points $Z' = \mathbf{F}(\mathbf{x}') \in \mathbf{Z}_{(n)}$.

Define $Q_{PCA}(Z')$ and $\mathbf{G}_{PCA}(\mathbf{x}')$ at OoS point $Z' = \mathbf{F}(\mathbf{x}') \notin \mathbf{Z}_{(n)}$ as follows. Let $\tilde{\mathbf{y}}(\mathbf{x})$ be arbitrary estimator for the value $\mathbf{f}(\mathbf{x})$ with convergence rate $O(n^{-(1+\eta)/(q+2)})$, $0 < \eta$, for example $\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{f}_{MLR}(\mathbf{x})$ for which $\eta = 1$, see (4); note than $\eta \leq 1$ in any case Genovese et al. (2012). In the future, the parameter γ will be chosen as satisfying the condition $0 < \gamma \leq \eta$.

Consider covariance matrix $\Sigma_n(\tilde{Z}'(\mathbf{x})|\mathbf{Z}_{(n)})$, which is calculated by formula (28) at the point $Z' = \tilde{Z}(\mathbf{x}') = \begin{pmatrix} \mathbf{x}' \\ \tilde{\mathbf{y}}(\mathbf{x}') \end{pmatrix}$ and gives the PCA-matrix $Q_{PCA}(\tilde{Z}(\mathbf{x}'))$. Using formula (38), introduce the quantity $\mathbf{G}_{PCA}(\mathbf{x}') = \mathbf{G}_{PCA}(\mathbf{x}'|\tilde{\mathbf{y}}(\mathbf{x}'))$, which approximates the $\mathbf{G}_{MLR}(\mathbf{x})$ with convergence rate $O(\varepsilon^{1+\eta})$. Therefore, the relation (61) yields

$$\mathbf{G}_{MLR}\left(\mathbf{x}\middle|\mathbf{Z}_{(n)/i} \cup \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix}\right) = \mathbf{G}_{PCA}(\mathbf{x}|\tilde{\mathbf{y}}(\mathbf{x})) - \tau(\mathbf{x}, \mathbf{y}) + O(\varepsilon^{2(1+\eta)}). \tag{62}$$

4. Conformal prediction algorithm

Based on approximations (50) and (62), we consider the approximations

$$\begin{aligned} b_i(\mathbf{x}, \mathbf{y}|\mathbf{Z}_{(n)/i} \cup (\mathbf{x}, \mathbf{y})^{\mathrm{T}}) &= \sum_{j=1, j\neq i}^{n} K(\mathbf{x}_j, \mathbf{x}_i) \times |\mathbf{y}_j - \mathbf{y}_i + \mathbf{G}_{PCA}(\mathbf{x}_j) \times (\mathbf{x}_j - \mathbf{x}_i)| \\ &+ K(\mathbf{x}, \mathbf{x}_i) \times |\mathbf{y} - \mathbf{y}_i + (\mathbf{G}_{PCA}(\mathbf{x}|\tilde{\mathbf{y}}(\mathbf{x})) - \tau(\mathbf{x}, \mathbf{y})) \times (\mathbf{x} - \mathbf{x}_i)| + O(\varepsilon^{2+\eta}) \\ &= b(\mathbf{x}_i, \mathbf{y}_i) + K(\mathbf{x}, \mathbf{x}_i) \times |\mathbf{y} - \mathbf{y}_i + \mathbf{G}_{PCA}(\mathbf{x}|\tilde{\mathbf{y}}(\mathbf{x})) \times (\mathbf{x} - \mathbf{x}_i) \\ &+ (\varepsilon \times \sigma_q)^{-2} \times (\mathbf{y} - \tilde{\mathbf{y}}(\mathbf{x})) \times [(\overline{\mathbf{x}} - \mathbf{x})^{\mathrm{T}} \times (\mathbf{x}_i - \mathbf{x})]| + O(\varepsilon^{2(1+\eta)}) \\ &= b(\mathbf{x}_i, \mathbf{y}_i) + K(\mathbf{x}, \mathbf{x}_i) \times |\mathbf{y} - \mathbf{y}_i + \mathbf{G}_{PCA}(\mathbf{x}|\tilde{\mathbf{y}}(\mathbf{x})) \times (\mathbf{x} - \mathbf{x}_i)| + O(\varepsilon^{2+\eta}) \end{aligned}$$

for the quantities $b_i(\mathbf{x}, \mathbf{y})$, $i = 1, 2, \ldots, n$ (10), here we use the fact that

$$K(\mathbf{x}, \mathbf{x}_i) \times (\mathbf{y} - \tilde{\mathbf{y}}(\mathbf{x})) \times [(\overline{\mathbf{x}} - \mathbf{x})^{\mathrm{T}} \times (\mathbf{x}_i - \mathbf{x})] = o(\varepsilon^{2+\eta}).$$

The PCA-matrices $\{\mathbf{G}_{PCA}(\mathbf{x}_j)\}$ at sample points $\{\mathbf{x}_j\}$ and $\mathbf{G}_{PCA}(\mathbf{x}|\tilde{\mathbf{y}}(\mathbf{x}))$ at OoS point $\mathbf{x}$ don't depend on MLR estimator and are defined by arbitrary chosen estimator $\tilde{\mathbf{y}}(\mathbf{x})$ with convergence rate $O(n^{-(1+\eta)/(q+2)})$, $0 < \eta \leq 1$.

Finally, we propose following nonconformity measures

$$\beta(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} K(\mathbf{x}_j, \mathbf{x}) \times |\mathbf{y}_j - \mathbf{y} + \mathbf{G}_{PCA}(\mathbf{x}_j - \mathbf{x})|,$$

$$\beta_i(\mathbf{x}, \mathbf{y}) = \beta(\mathbf{x}_i, \mathbf{y}_i) + K(\mathbf{x}, \mathbf{x}_i) \times |\mathbf{y} - \mathbf{y}_i + \mathbf{G}_{PCA}(\mathbf{x}|\tilde{\mathbf{y}}(\mathbf{x})) \times (\mathbf{x} - \mathbf{x}_i)|.$$

Following Vovk et al. (2005); Shafer and Vovk (2008); Papadopoulos et al. (2014), prediction region $\Gamma_\alpha(\mathbf{x})$ for unknown output $\mathbf{y} = \mathbf{f}(\mathbf{x})$ at OoS input point $\mathbf{x}$ with given confidence level α based on the MLR-estimators $\mathbf{G}_{\mathbf{f}}(\mathbf{x})$ is

$$\Gamma_\alpha(\mathbf{x}) = \left\{\mathbf{y} \in W_{C,\gamma}(\mathbf{x}, \tilde{\mathbf{y}}(\mathbf{x})) : \ \frac{1}{n+1} + \frac{1}{n+1} \sum_{i=1}^{n} \mathrm{I}\left[\beta_i(\mathbf{x}, \mathbf{y}) \geq \beta(\mathbf{x}, \mathbf{y})\right] > \alpha\right\}$$

in which the set $W_{C,\gamma}(\tilde{\mathbf{y}}(\mathbf{x}))$ is defined in (26) with chosen values C and $\gamma, 0 < \gamma \leq \eta$.

5. Conclusion

Standard multi-output regression task with multivariate inputs is considered. A new nonconformity measure for this regression task and using an analog of Bregman distance is proposed. This allowed to construct prediction region for unknown output $\mathbf{y} = \mathbf{f}(\mathbf{x})$ at OoS input point $\mathbf{x}$ with given confidence level α.

Acknowledgments

This work is supported by the Russian Science Foundation (project 14-50-00150).

References

S. Amari and H Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical monographs*. Oxford University Press, 2000.

Shun-Ichi Amari. Natural gradient works efficiently in learning. *Neural Comput.*, 10(2): 251–276, February 1998. ISSN 0899-7667. doi: 10.1162/089976698300017746. URL http://dx.doi.org/10.1162/089976698300017746.

A. Banerjee, X. Guo, and H. Wang. On the optimality of conditional expectation as a bregman predictor. *IEEE Trans. Inf. Theor.*, 51(7):2664–2669, July 2005a. ISSN 0018-9448. doi: 10.1109/TIT.2005.850145. URL http://dx.doi.org/10.1109/TIT.2005.850145.

Arindam Banerjee, Srujana Merugu, Inderjit S. Dhillon, and Joydeep Ghosh. Clustering with bregman divergences. *J. Mach. Learn. Res.*, 6:1705–1749, December 2005b. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1046920.1194902.

Mikhail Belyaev, Evgeny Burnaev, and Yermek Kapushev. Gaussian process regression for structured data sets. In Alexander Gammerman, Vladimir Vovk, and Harris Papadopoulos, editors, *Statistical Learning and Data Sciences*, pages 106–115, Cham, 2015a. Springer International Publishing. ISBN 978-3-319-17091-6.

Mikhail Belyaev, Evgeny Burnaev, and Yermek Kapushev. Gaussian process regression for structured data sets. In Alexander Gammerman, Vladimir Vovk, and Harris Papadopoulos, editors, *Statistical Learning and Data Sciences*, pages 106–115, Cham, 2015b. Springer International Publishing. ISBN 978-3-319-17091-6.

Mikhail Belyaev, Evgeny Burnaev, Ermek Kapushev, Maxim Panov, Pavel Prikhodko, Dmitry Vetrov, and Dmitry Yarotsky. Gtapprox: surrogate modeling for industrial design. *CoRR*, abs/1609.01088, 2016. URL http://arxiv.org/abs/1609.01088.

Alexander V. Bernstein and Alexander P. Kuleshov. Tangent bundle manifold learning via grassmann&stiefel eigenmaps. *CoRR*, abs/1212.6031, 2012. URL http://arxiv.org/abs/1212.6031.

Alexander V. Bernstein and Alexander P. Kuleshov. Manifold learning: Generalization ability and tangent proximity. *Int. J. Software and Informatics*, 7(3):359–390, 2013. URL http://www.ijsi.org/ch/reader/view_abstract.aspx?file_no=i165.

Alexander V. Bernstein, Alexander P. Kuleshov, and Yury Yanovich. Statistical learning via manifold learning. In *14th IEEE International Conference on Machine Learning and Applications, ICMLA 2015, Miami, FL, USA, December 9-11, 2015*, pages 64–69, 2015. doi: 10.1109/ICMLA.2015.26. URL https://doi.org/10.1109/ICMLA.2015.26.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

L.M. Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *Zh. Vychisl. Mat. Mat. Fiz.*, 7:620–631, 1967. ISSN 0044-4669.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

E. V. Burnaev, M. E. Panov, and A. A. Zaytsev. Regression on the basis of nonstationary gaussian processes with bayesian regularization. *Journal of Communications Technology and Electronics*, 61(6):661–671, Jun 2016. ISSN 1555-6557. doi: 10.1134/S1064226916060061. URL https://doi.org/10.1134/S1064226916060061.

E.V. Burnaev, A.A. Zaytsev, and V.G. Spokoiny. The Bernstein-von Mises theorem for regression based on Gaussian processes. *Russ. Math. Surv.*, 68(5):954–956, 2013. ISSN 0036-0279; 1468-4829/e. doi: 10.1070/RM2013v068n05ABEH004863.

Evgeny Burnaev and Ivan Nazarov. Conformalized kernel ridge regression. In *15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, Anaheim, CA, USA, December 18-20, 2016*, pages 45–52, 2016. doi: 10.1109/ICMLA.2016.0017. URL https://doi.org/10.1109/ICMLA.2016.0017.

Evgeny Burnaev and Maxim Panov. Adaptive design of experiments based on gaussian processes. In Alexander Gammerman, Vladimir Vovk, and Harris Papadopoulos, editors, *Statistical Learning and Data Sciences*, pages 116–125, Cham, 2015. Springer International Publishing. ISBN 978-3-319-17091-6.

Evgeny Burnaev and Vladimir Vovk. Efficiency of conformalized ridge regression. *CoRR*, abs/1404.2083, 2014. URL http://arxiv.org/abs/1404.2083.

L. Cayton. Algorithms for manifold learning. *Univ. of California at San Diego Tech. Rep*, 2005.

Li Deng and Dong Yu. Deep learning: Methods and applications. *Found. Trends Signal Process.*, 7(3–4):197–387, June 2014. ISSN 1932-8346. doi: 10.1561/2000000039. URL http://dx.doi.org/10.1561/2000000039.

Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 2000.

Christopher R. Genovese, Marco Perone-Pacifico, Isabella Verdinelli, and Larry Wasserman. Minimax manifold estimation. *J. Mach. Learn. Res.*, 13:1263–1291, May 2012. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=2188385.2343687.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, 2009.

Xiaoming Huo, Xuelei (Sherry) Ni, and Andrew K. Smith. *A Survey of Manifold-Based Learning Methods*, pages 691–745. WORLD SCIENTIFIC, 2011.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014. ISBN 1461471370, 9781461471370.

Jurgen Jost. Riemannian geometry and geometric analysis. 2002.

D. N. Kaslovsky and F. G. Meyer. Non-Asymptotic Analysis of Tangent Space Perturbation. *ArXiv e-prints*, November 2011.

Alexander P. Kuleshov and Alexander Bernstein. Nonlinear multi-output regression on unknown input manifold. *Ann. Math. Artif. Intell.*, 81(1-2):209–240, 2017. doi: 10.1007/s10472-017-9551-0. URL https://doi.org/10.1007/s10472-017-9551-0.

Alexander P. Kuleshov and Alexander V. Bernstein. Manifold learning in data mining tasks. In Petra Perner, editor, *Machine Learning and Data Mining in Pattern Recognition - 10th International Conference, MLDM 2014, St. Petersburg, Russia, July 21-24, 2014. Proceedings*, volume 8556 of *Lecture Notes in Computer Science*, pages 119–133. Springer, 2014a. doi: 10.1007/978-3-319-08979-9_10. URL https://doi.org/10.1007/978-3-319-08979-9_10.

Alexander P. Kuleshov and Alexander V. Bernstein. Data-based manifold reconstruction via tangent bundle manifold learning. In *ICML-2014, Workshop "Topological Methods for Machine Learning"*, pages 1–4, 2014b. URL http://topology.cs.wisc.edu/KuleshovBernstein.pdf.

Alexander P. Kuleshov and Alexander V. Bernstein. Extended regression on manifolds estimation. In *Conformal and Probabilistic Prediction with Applications - 5th International Symposium, COPA 2016, Madrid, Spain, April 20-22, 2016, Proceedings*, pages 208–228, 2016. doi: 10.1007/978-3-319-33395-3_15. URL https://doi.org/10.1007/978-3-319-33395-3_15.

Jeffrey M. Lee. *Manifolds and Differential Geometry*, volume 107. American Mathematical Society, 2009.

Elizaveta Levina and Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. In L. K. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information*

Processing Systems 17, pages 777–784. MIT Press, 2005. URL http://papers.nips.cc/paper/2577-maximum-likelihood-estimation-of-intrinsic-dimension.pdf.

Clive Loader. *Local regression and likelihood.* New York: Springer-Verlag, 1999. ISBN 0-387-9877.

Yunqian Ma and Yun Fu. *Manifold Learning Theory and Applications.* CRC Press, Inc., 2011.

Harris Papadopoulos, Vladimir Vovk, and Alex J. Gammerman. Regression conformal prediction with nearest neighbours. *CoRR*, abs/1401.3880, 2014. URL http://arxiv.org/abs/1401.3880.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning).* The MIT Press, 2005. ISBN 026218253X.

Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *J. Mach. Learn. Res.*, 9:371–421, June 2008. ISSN 1532-4435. URL http://dl.acm.org/citation.cfm?id=1390681.1390693.

A. Singer and H.T. Wu. Vector diffusion maps and the connection laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144. doi: 10.1002/cpa.21395. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21395.

H. Tyagi, E. Vural, and P. Frossard. Tangent space estimation for smooth embeddings of Riemannian manifolds. *ArXiv e-prints*, August 2012.

Vladimir N. Vapnik. *Statistical Learning Theory.* Wiley-Interscience, 1998.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World.* Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2005. ISBN 0387001522.

Yury Yanovich. Asymptotic properties of local sampling on manifold. *Journal of Mathematics and Statistics*, 12(3):157–175, 2016.

A. A. Zaitsev, E. V. Burnaev, and V. G. Spokoiny. Properties of the posterior distribution of a regression model based on gaussian random fields. *Automation and Remote Control*, 74(10):1645–1655, Oct 2013. ISSN 1608-3032. doi: 10.1134/S0005117913100056. URL https://doi.org/10.1134/S0005117913100056.