

The Mutual Nearest Neighbor Method in Functional Nonparametric Regression

Xingyu Chen^{1,*}, Dirong Chen²

¹School of Mathematics and System Science, Beihang University, Beijing, P. R. China

²School of Mathematics and Computer Sciences, Wuhan Textile University, Wuhan, P. R. China

Email address:

chengxingyu@buaa.edu.cn (Xingyu Chen)

*Corresponding author

To cite this article:

Xingyu Chen, Dirong Chen. The Mutual Nearest Neighbor Method in Functional Nonparametric Regression. *Science Journal of Applied Mathematics and Statistics*. Vol. 6, No. 3, 2018, pp. 81-89. doi: 10.11648/j.sjams.20180603.13

Received: June 5, 2018; **Accepted:** July 12, 2018; **Published:** July 19, 2018

Abstract: In recent decades, functional data have become a commonly encountered type of data. Its ideal units of observation are functions defined on some continuous domain and the observed data are sampled on a discrete grid. An important problem in functional data analysis is how to fit regression models with scalar responses and functional predictors (scalar-on-function regression). This paper focuses on the nonparametric approaches to this problem. First there is a review of the classical k-nearest neighbors (kNN) method for functional regression. Then the mutual nearest neighbors (MNN) method, which is a variant of kNN method, is applied to functional regression. Compared with the classical kNN approach, the MNN method takes use of the concept of mutual nearest neighbors to construct regression model and the pseudo nearest neighbors will not be taken into account during the prediction process. In addition, any nonparametric method in the functional data cases is affected by the curse of infinite dimensionality. To prevent this curse, it is legitimate to measure the proximity between two curves via a semi-metric. The effectiveness of MNN method is illustrated by comparing the predictive power of MNN method with kNN method first on the simulated datasets and then on a real chemometrical example. The comparative experimental analyses show that MNN method preserves the main merits inherent in kNN method and achieves better performances with proper proximity measures.

Keywords: Functional Data, Nonparametric Estimation, Mutual Nearest Neighbors Estimator, Semi-Metric

1. Introduction

Regression is a well-known statistical problem, which consists in studying the link between the explanatory variable and the response variable. The regression function, which is most commonly the conditional expectation of the explanatory variable, is to be estimated in order to predict the response variable given a new value for the explanatory variable.

In the classical statistical setting, each observation X_i is supposed to be a collection of numerical measurements represented by a p -dimensional vector. In practice, observations are observed at discrete sampling points. Thanks to technological progress, instants of measure come closer so grids are more and more fine. These practical improvements have naturally cast the regression problem into the general class of functional data analysis, where the

observation contains at least a functional object rather than standard vectors.

Functional data are intrinsically infinite dimensional. On the one hand, the infinite dimensional structure of the functional data is a rich source of information. On the other hand, the high intrinsic dimensionality of these data poses challenges both for theory and computation. The book of Ramsay and Silverman [1] provides a comprehensive introduction to the functional data analysis. Nonparametric ideas have been popularized by the book of Ferraty and Vieu [2]. A brief discussion of functional data analysis is presented by Wang et al. [3] and an introduction to recent advances in infinite dimensional statistics is presented by Goia and Vieu [4].

Morris [5] provides an introduction to the regression with functional data. According to the role played by the functional data in each model, functional regression models are classified into three categories: scalar-on-function

regression (scalar responses and functional predictors), function-on-scalar regression (functional responses and scalar predictors) and function-on-function regression (functional responses and functional predictors). The scalar-on-function regression has been applied in many domains and Reiss et al. [6] reviews some approaches to this regression problem. This article focuses on the nonparametric approaches to scalar-on-function regression.

The k-nearest neighbor (kNN) is one of the most famous nonparametric method. The first works were presented by Royall [7] and Stone [8], and the monograph by Györfi et al. [9] proposes a wide presentation on kNN ideas in nonparametric finite dimensional statistics. Starting with Laloë [10] and Burba et al. [11], there are some advances in kNN nonparametric functional data analysis, including the construction of the kNN kernel estimator of regression operator and some asymptotic results with convergence rates.

The mutual nearest neighbors (MNN) method is a variant of kNN method. It takes use of the concept of mutual nearest neighbors rather than k-nearest neighbors to construct regression model. The term of mutual nearest neighbors seems to date back to Chidanand Gowda and Krihna [12] in the context of clustering. Liu et al. [13] apply mutual nearest neighbors to classification purposes and the experimental results show that the classification performance achieved by MNN method is better than the classical kNN method. Guyader et al. [14] investigates the theoretical properties of MNN method for the finite dimensional regression case.

This paper proposes the mutual nearest neighbors method in nonparametric functional regression setting. The organization of the paper is as follows. Section 2 presents MNN estimators based on a functional regression problem and some proximity measures. Section 3 shows the experimental results on the simulated datasets and then Section 4 shows that MNN method achieves a better performance on a real chemometrical example.

2. MNN Method and Proximity Measures

2.1. Models and Estimators

A random element is said to be functional if it takes its values in an infinite dimensional space. Many random elements, such as vectors of random curves, random surfaces and random fields of any dimension, enter this general definition. In this paper, functional data typically consist of a random sample of independent real-valued functions, $X_1(t), \dots, X_n(t)$, on a compact interval T on a real line. Such data have also been termed curve data.

The regression problem now is about linking the response Y to a whole curve observed on a domain T

$$X \triangleq \{X(t): t \in T\}.$$

The random curve X is assumed to belong to a functional space \mathcal{F} . Therefore, the functional regression model we consider is

$$Y = r(X) + \varepsilon,$$

with $Y \in \mathbb{R}$ and ε a scalar random disturbance such that $E(\varepsilon|X) = 0$ almost surely. And now $r(\cdot)$ is an operator from \mathcal{F} to \mathbb{R} satisfying some necessary regularity conditions.

Let $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ be a sample of the functional random variable X and response variable Y valued in $\mathcal{F} \times \mathbb{R}$. Given a new observation curve x , our goal is to estimate the regression function $r(x) = E[Y|X = x]$ using the data \mathcal{D}_n .

For a fixed x , reorder the data \mathcal{D}_n according to the increasing values of $d(X_i, x)$, where $d(\cdot, \cdot)$ is a function that defines the proximity between each pair of elements of a set. The reordered data sequence is denoted by

$$(X_{(1,n)}(x), Y_{(1,n)}(x)), \dots, (X_{(n,n)}(x), Y_{(n,n)}(x)).$$

If no confusion is possible, it can also be denoted by

$$(X_{(1)}(x), Y_{(1)}(x)), \dots, (X_{(n)}(x), Y_{(n)}(x)).$$

$X_{(k)}(x)$ is called the k -th nearest neighbor of x . If $d(X_i, x) = d(X_j, x)$, then we have a tie. The tie breaking is done by indices, i.e., X_i might be declared "closer" if $i < j$.

Let us denote $\mathcal{N}_k(x)$ the set of the k nearest neighbors of x in \mathcal{D}_n ,

$$\mathcal{N}_k(x) = \{X_{(1)}(x), \dots, X_{(k)}(x)\}$$

For a fixed $x \in \mathcal{F}$, the kNN estimator can be written as

$$\hat{r}_{knn}(x) = \sum_{i: X_i \in \mathcal{N}_k(x)} Y_i.$$

From an optimal perspective, it is not a good idea to fix a globally optimal value of k for kNN in some real applications. For example, although an outlier is far from other curves, it may still treat some other curves as its nearest neighbors. Such pseudo neighbors are often found in the classical k-nearest neighbors. If they are used to predict the responses of new curves, the final results are doubttable in some way. Therefore, it would be better to exclude the pseudo neighbors from constructing regression models.

Given two parameters k and l , let us denote $\mathcal{N}_k(x)$ the set of the k nearest neighbors of x in \mathcal{D}_n , $\mathcal{N}_l(x)$ the set of the l nearest neighbors of X_i in $(\mathcal{D}_n \setminus \{X_i\}) \cup \{x\}$. The set of the mutual nearest neighbors of x is

$$\mathcal{M}_{k,l}(x) = \{X_i: X_i \in \mathcal{N}_k(x), x \in \mathcal{N}_l(X_i)\}.$$

Denoting $M_{k,l}(x) = |\mathcal{M}_{k,l}(x)|$ the number of mutual nearest neighbors of x . For the sake of simplicity, k and l are usually assigned to the same value k . Thus, $\mathcal{M}_{k,l}(x)$ and $M_{k,l}(x)$ can be denoted as $\mathcal{M}_k(x)$ and $M_k(x)$ respectively. And then $M_k(x)$ is a random variable taking values in $\{0, 1, \dots, k\}$.

An observation curve X_1 is a mutual nearest neighbor of X_2 , if and only if X_1 appears in the k -nearest neighbors list

of X_2 and X_2 appears in the k -nearest neighbors list of X_1 at the same time. Obviously, X_2 is also a mutual nearest neighbor of X_1 . It means that mutual nearest neighbors have symmetric property contrarily to the k -nearest neighbors.

If only the mutual nearest neighbors are used to construct the regression model, MNN regression estimate is defined as follows

$$\hat{f}_{mnn}(x) = \frac{1}{M_k(x)} \sum_{i: X_i \in \mathcal{M}_k(x)} Y_i$$

with the convention that $0/0 = 0$. It is an extension of kNN functional regression with mutual nearest neighbors. Because the pseudo neighbors of x have been eliminated and the rest are closer to itself, the MNN regression method is more reasonable than kNN regression method.

2.2. Proximity Measures

Generally, the performance of both nearest neighbors method heavily relies on the following three factors: the sample size n , the selection of proximity measure $d(\cdot, \cdot)$ and the number of neighbors k .

In the classical multivariate regression model, X is a p -dimensional vector of continuous covariates and $r(\cdot)$ is a smooth function from \mathbb{R}^p to \mathbb{R} . In this case, the proximity measure $d(\cdot, \cdot)$ which is used to measure the distance between observations is the standard Euclidean metric of \mathbb{R}^p . Thus, the distance between X_i and X_j is

$$d(X_i, X_j) = \sqrt{(x_{i1} - x_{j1})^2 + \dots + (x_{ip} - x_{jp})^2}$$

where $X_i = (x_{i1}, \dots, x_{ip})^T$ and $X_j = (x_{j1}, \dots, x_{jp})^T$.

In the functional cases, the natural distance between elements are computed by means of the classical L_2 -metric, a generalization of the Euclidean metric, which is defined for all observed curves X_i and X_j by

$$d(X_i, X_j) = \sqrt{\int [X_i(t) - X_j(t)]^2 dt}. \quad (1)$$

But sometimes, L_2 -metric may lead to terrible results in nonparametric functional cases.

The curse of dimensionality is strongly linked with the sparseness of data in a high-dimensional space. The concentration function $\varphi(h)$ measures how densely packed are the considered elements of \mathcal{F} in an infinite dimensional ball of radius h . For a fixed $x \in \mathcal{F}$, the value of the concentration function is estimated by

$$\hat{\varphi}_x(h) = \frac{1}{n} \sum_{i=1}^n 1_{B(x,h)}(X_i), \quad (2)$$

where h is a fixed bandwidth, the symbol 1 is an indicator function and $B(x, h) = \{x': d(x, x') < h\}$. Geenens [15] conjectures that the concentration function $\varphi(h)$ typically decreases to 0 exponentially quickly as $h \rightarrow 0$ and the convergence of \hat{f} toward r has unacceptable logarithmic rate in the functional space of interest, which is called as the curse

of infinite dimensionality. Concretely, for a fixed x , there are too few X_i close enough to x in the sense of L_2 -metric even when the sample size grows to infinity. And hence it is not reliable to make the estimation of $r(x)$ using nearest neighbors.

On the other hand, the L_2 -metric just considers functions as infinite-dimensional vectors for it is the strict generalization of the Euclidean metric to functions. If the actual values taken by the functions are important for the problem, the L_2 -metric may be a good choice. But if only the shape of the functions is related to the response Y , the L_2 -metric may lead to bad conclusions. It can be explained that the L_2 -metric is probably too exacting to capture some of the features proper to functions, such as their general appearance, the way they vary over short or long range, etc.

Results of Ferraty and Vieu [2] show that it is wise to choose other looser proximity measures in place of L_2 -metric. The new proximity measures are expected to have the following two characteristics. First, it results in a new concentration function that decreases less rapidly as $h \rightarrow 0$. Therefore, more curves are close to each other so that more observations can contribute to the computation of $r(x)$. Second, it leaves out the irrelevant characteristics of the curves and highlights the particular characteristics of the curves that we know to be directly related to the response in the analysis.

It is a good idea to measure the proximity between two elements of the infinite dimensional functional space via a semi-metric, which is a metric except that $d(x, y) = 0 \nRightarrow x = y$.

For instance, consider the following family of semi-metrics:

$$d_q(X_i, X_j) = \sqrt{\int [X_i^{(q)}(t) - X_j^{(q)}(t)]^2 dt}, \quad (3)$$

where q takes values in $\{0, 1, 2, \dots\}$. For any q -times differentiable real function x , $x^{(q)}(t)$ denotes the q -th derivative of x . The main effect produced by the differentiating operator is to highlight some ranges of observed curves with large variations. The following example illustrated this effect.

Consider the three functions $x(t)$ (black line), $y(t)$ (blue dashed) and $z(t)$ (red twodash) presented in Figure 1. Intuitively, the general appearance of $x(t)$ and $y(t)$ is almost identical while $x(t)$ and $z(t)$ are different. Based on the L_2 -metric, the distance between $x(t)$ and $z(t)$ is smaller so that $z(t)$ becomes the nearest neighbor of $x(t)$. This may lead to a bad conclusion if the response is mainly related to the trend of curves, as $y(t)$ is more helpful to the prediction procedure of $x(t)$. In this case, the problem can be solved by the semi-metric based on the first derivatives:

$$d_1(x, y) = \sqrt{\int [x'(t) - y'(t)]^2 dt}.$$

This looser proximity measure cancels out the vertical shift and care more about the shape of the curves. With this semi-metric $d_1(\cdot, \cdot)$, $y(t)$ becomes the nearest neighbor of $x(t)$ and lead to a better prediction performance.

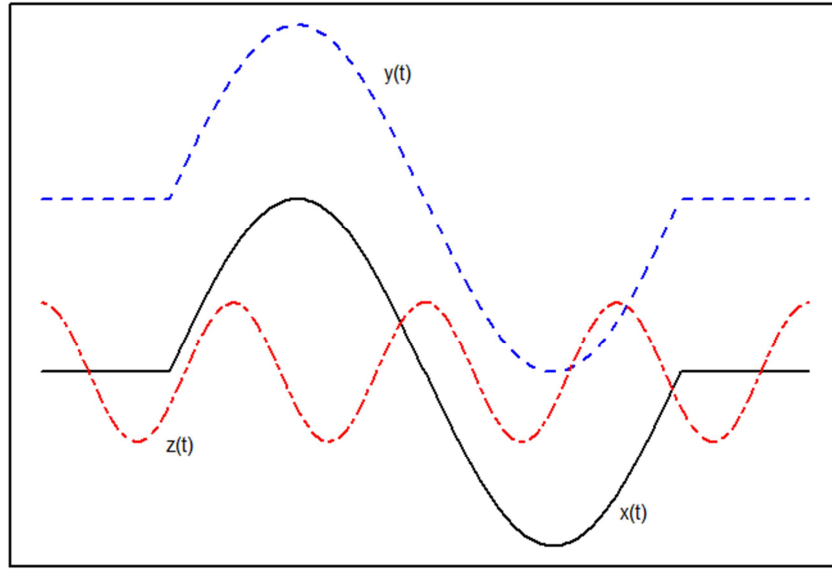


Figure 1. Three curves: $x(t)$ (black line), $y(t)$ (blue dashed) and $z(t)$ (red two dash).

3. Simulation Experiments

In order to demonstrate the effectiveness of the MNN method, this section shows that how the MNN procedures work in a simple simulated finite sample situation. As mentioned in the previous section, the MNN method is an extension of kNN by using mutual nearest neighbors. Thus a comparison between the classical kNN method and MNN method is designed in this section.

We simulate $n = 300$ pairs $\{(X_i, Y_i)\}_{i=1}^n$. The sampled functional explanatory variables are generated as follows, for each $i \in \{1, 2, \dots, 300\}$:

$$X_i \triangleq X_i(t) = a_i \cos(2t), t \in [0, \pi]$$

where

$$a_i \sim N(0, 1) \text{ for } i = 1, 2, \dots, 150,$$

$$a_i \sim N(3, \sigma^2) \text{ for } i = 151, \dots, 300.$$

Different values for σ^2 will create two groups of curves inside the dataset with concentration being different from one group to the other one. In this simulation study, two extreme cases are considered, the most homogeneous one when $\sigma^2 = 1$ and the most heterogeneous one when $\sigma^2 = 0.1$. The curves X_1, \dots, X_n are observed on the same grid generated from 100 equispaced measurements in $[0, \pi]$ and displayed in Figure 2. (Note that these cases are the same as the cases in Burba et al. [7].)

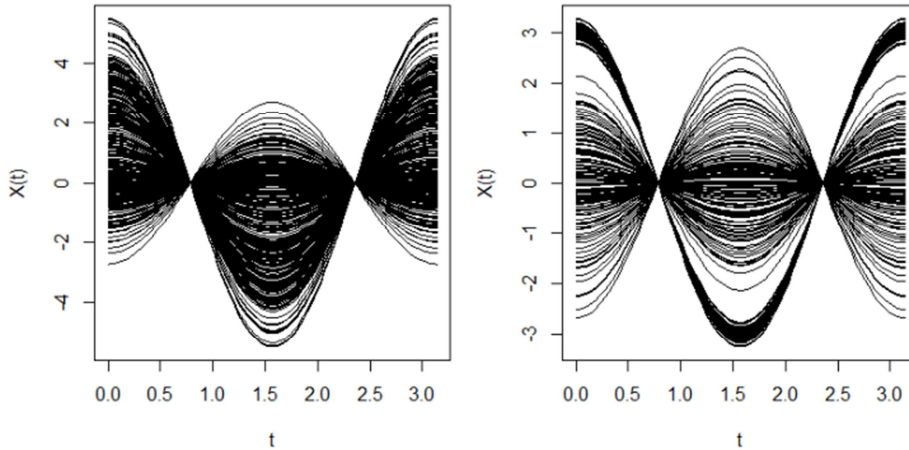


Figure 2. The simulated curves with $\sigma^2 = 1$ (left panel) and the simulated curves with $\sigma^2 = 0.1$ (right panel).

Here the notions of homogeneity and heterogeneity will refer to the concentration of the data. The concentration function $\varphi(h)$, which can be estimated the values by (2), illustrates the difference between the two datasets. Usually, the bandwidth h is obtained by a cross-validation procedure but other values of h have similar results in fact. Here 0.7 is

chosen as the value of h and the classical L_2 -metric (1) is used to measure the proximity between two curves. The left plot in Figure 3 shows homogeneity since each curve has roughly the same number of neighbors. And the right plot in Figure 3 shows high heterogeneity since both groups of curves have very different numbers of neighbors.

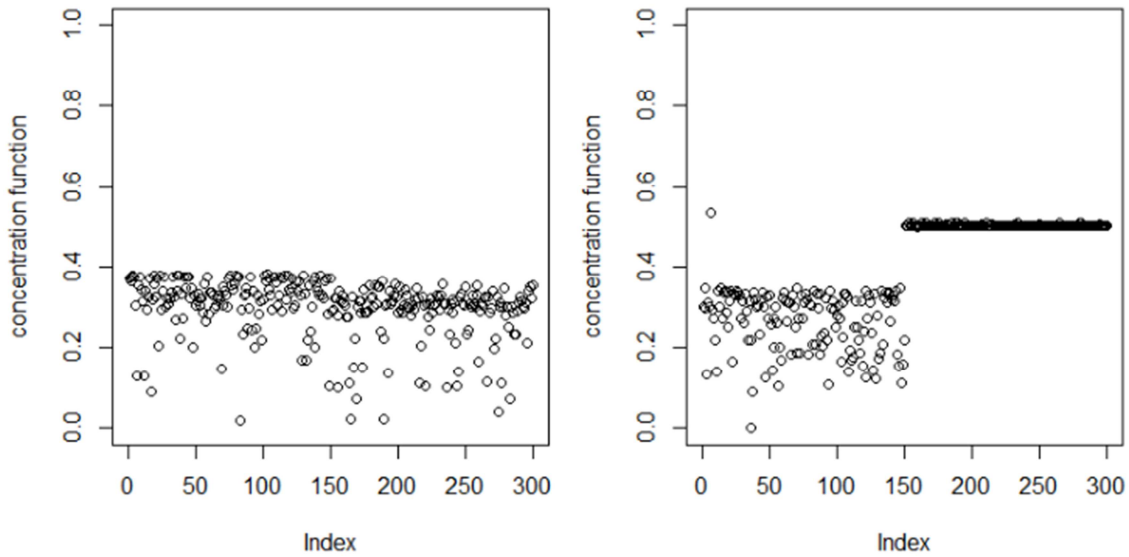


Figure 3. Values of concentration function with $h = 0.7$ (left: $\sigma^2 = 1$; right: $\sigma^2 = 0.1$).

The aim is to study the following functional nonparametric regression model:

$$Y_i = r(X_i) + \varepsilon_i,$$

where the ε_i 's are generated independently according to a $N(0, 0.05)$ distribution and are assumed to be independent of X_i for all $i \in \{1, 2, \dots, n\}$. We will observe the behavior of the two methods (MNN method and kNN method) in these two extreme cases. The nonparametric regression operator $r(x)$ is

$$r(X_i) = a_i^2, i = 1, \dots, n.$$

In both cases 150 curves will be randomly taken to construct the training sample S_{train} and the other 150 will constitute the testing sample S_{test} . Both kNN-based estimators and MNN-based estimators depend on the quantity of neighbors k which must be selected from the training sample. And the testing sample is used to measure the accuracy of the predictions. The proximity is measured by the classical L_2 -metric.

The optimal number of neighbors k_{opt} is obtained by minimizing a global cross-validation criterion on the training sample, that is:

$$k_{opt} = \arg \min_k CV(k)$$

where

$$CV(k) = \sum_{i=1}^n [Y_i - \hat{r}^{(-i)}(X_i)]^2$$

with

$$\hat{r}^{(-i)}(x) = \frac{1}{k} \sum_{j: j \in \mathcal{N}_k(x), j \neq i} Y_j$$

for kNN-based estimator and

$$\hat{r}^{(-i)}(x) = \frac{1}{M_k(x)} \sum_{j: j \in \mathcal{M}_k(x)} Y_j$$

for MNN-based estimator. Noting that the optimal number of neighbors k_{opt} or these two methods are different in most times.

Some results of predictions are presented here. Note that MSEP is the mean square error of prediction, that is

$$MSEP = \frac{1}{\#(S_{test})} \sum_{i \in S_{test}} (Y_i - \hat{Y}_i)^2,$$

where $\#(S_{test})$ denotes the quantity of the testing sample and \hat{Y}_i denotes a prediction of Y_i . Here $\#(S_{test})$ is 150.

Figure 4 shows the prediction for the most homogeneous case ($\sigma^2 = 1$), while Figure 5 shows the prediction for the most heterogeneous case ($\sigma^2 = 0.1$). An advantage of kNN method is that it takes into account the local structure of the data and gives good predictions for the heterogeneous case, as presented by Burba et al. The proposed MNN method can also do this and is more efficient than kNN method for both cases. For the homogeneous dataset, the MSEP for the MNN-based estimator is 0.1470 against 1.0508 for the kNN-based estimator. And the MNN method is still better in the heterogeneous situation with a MSEP of 0.0155 against 0.0845 for the kNN method.

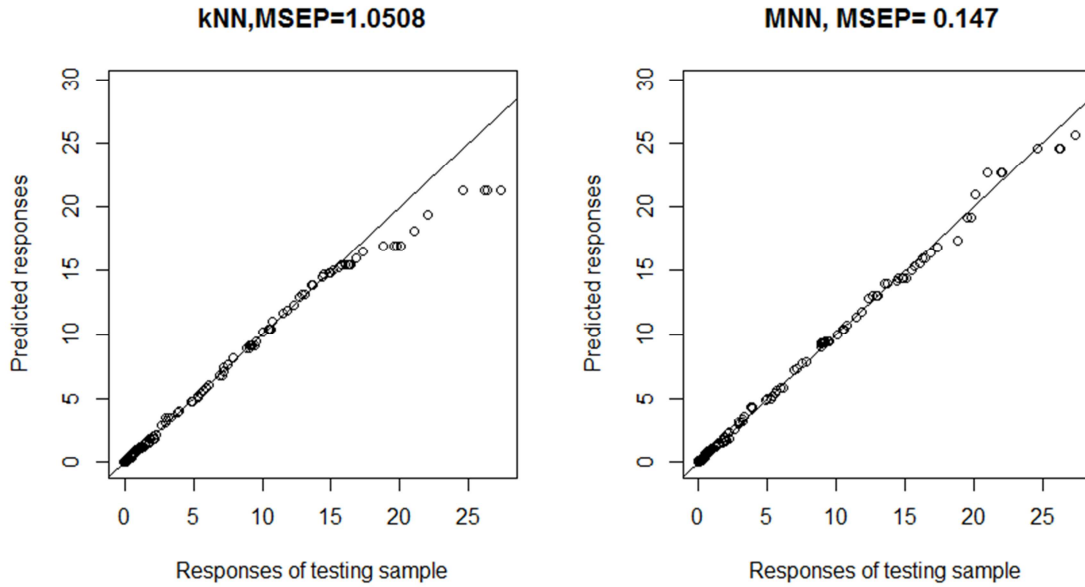


Figure 4. Knn method vs. mnn method in the most homogeneous case.

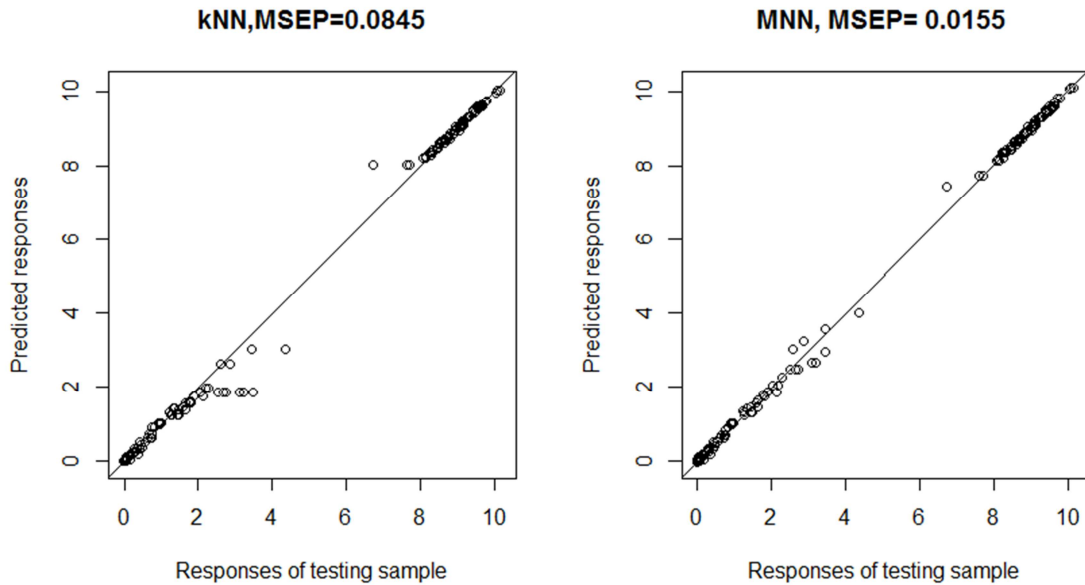


Figure 5. Knn method vs. mnn method in the most heterogeneous case.

It is obvious that for the kNN method, there is a deviation between \hat{Y}_i and Y_i when the true value locates at the boundary of the range. This is due to the fact that the k nearest training samples are used to predict the response of the curve we concern in kNN method. When the curve, represented by x_0 , locates at the boundary, the majority of the training samples are on the same side of x_0 . If the k -th nearest training sample is far away from x_0 , the deviation could not be avoided.

But for the MNN method, the deviation is not obvious. The MNN method takes use of the concept of mutual nearest neighbors rather than k nearest neighbors of x_0 to predict its values. The sample curves which are far away from x_0 may be excluded so that they will not be taken as determinant conditions. Thus, the final prediction results for MNN method

don't have obvious deviation.

Replicate the experiment $M = 100$ times and then M samples $\{(X_i, Y_i)\}_{i=1}^n$ are drawn. Each of the M simulated datasets are randomly split into two sample: a training sample S_{train} contained 150 curves and a test sample S_{test} contained other 150 curves. From each training sample and or both kNN and MNN methods, the tuning parameter k is selected by means of the global cross-validation procedure and $Y_i (i \in S_{test})$ is predicted. The accuracy of predictions of Y_i is measured by means of the MESP. Figure 6 displays boxplots of the corresponding M values of the MESP obtained from each of the M replicates of the experiment. From the predictive power point of view, the proposed MNN method preserves the main merits inherent in kNN method and is more effective.

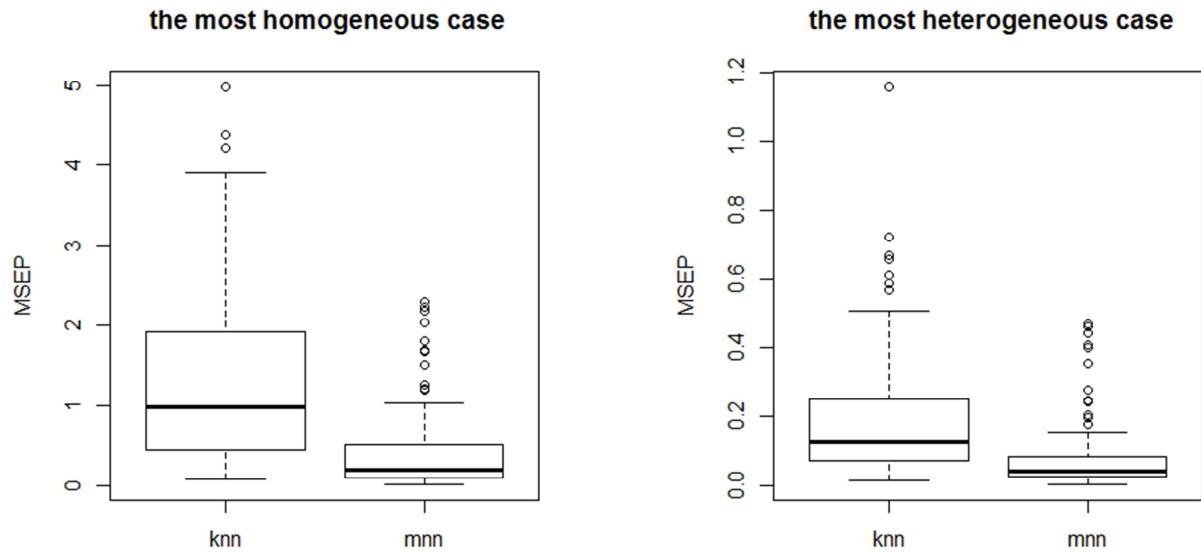


Figure 6. Mean square errors of prediction from the $M = 100$ replicates of the experiment. Left panel: $\sigma^2 = 1$ (the most homogeneous case). Right panel: $\sigma^2 = 0.1$ (the most heterogeneous case).

4. A Real Dataset Application

The aim of this section is to compare the predictive power of the MNN method and kNN method on a real dataset. We focus here on a quality control problem in the food industry. The dataset concerns a sample of finely chopped meat and is a part of Tecator dataset, which can be found at <http://lib.stat.cmu.edu/datasets/tecator>.

The dataset consists in $N = 215$ observations. For each meat sample i , the data consists of a 100 channel spectrum of absorbance X_i and the contents of fat Y_i . The absorbance is $-\ln 10$ of the transmittance measured by the spectrometer and the data were recorded on a Tecator Infracore Food and Feed Analyzer working in the wavelength range 850 – 1050 nm by the near-infrared (NIR) transmission principle. Because of the fineness of the grid, each sample can be considered as a

continuous curve although it appears clearly as a discretized curve. The spectrometric curves are presented in the left panel of Figure 7. The content of fat, measured in percent, is determined by analytic chemistry. The task is to predict the fat content of a meat sample on the basis of its near infrared absorbance spectrum.

To compare the predictive power of each functional regression method (kNN and MNN), this sample of size 215 was splitted into a learning sample of size 120 and a testing sample of size 95. First, consider the partition given by $S_{train} = \{1, 2, \dots, 120\}$ and $S_{test} = \{121, \dots, 215\}$ to obtain a proper proximity measure. The L_2 -metric, quantifying in a sense the area between the curves, focuses on the vertical shifts among curves and thus lead to poor predictions. The MSEP with the L_2 -metric for MNN method is 138.57.

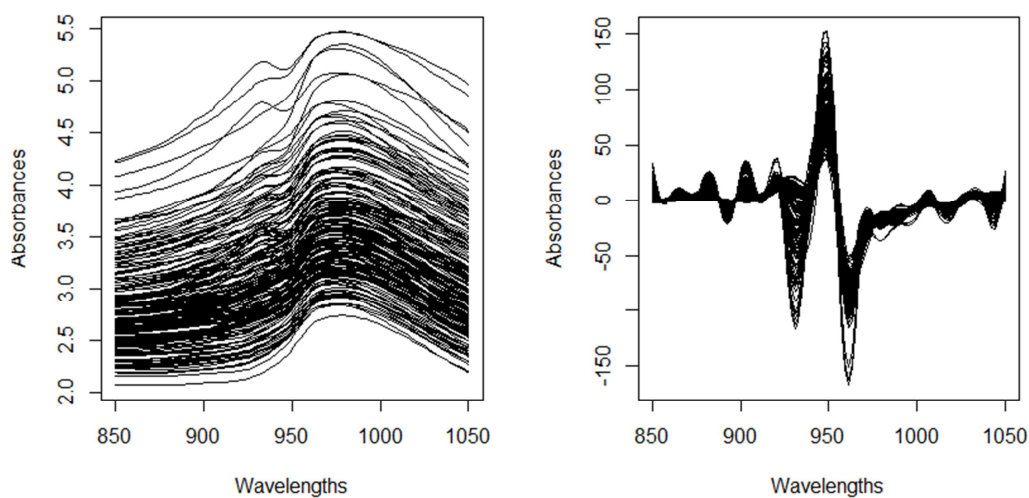


Figure 7. Absorbance curves (left panel) and their second derivatives (right panel).

Since the smoothness of the curves, we consider the family of semi-metrics based on the derivatives, $\{d_q(\cdot, \cdot), q =$

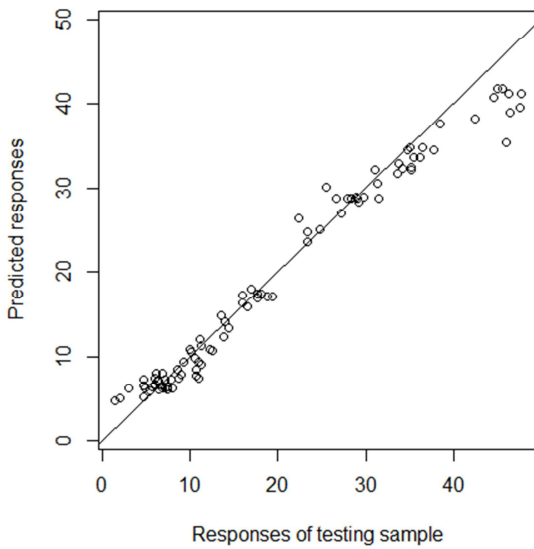
$1, 2, 3, 4\}$, where $d_q(\cdot, \cdot)$ is represented as (3). Table 1 reports the values of both the MSEP and the selected semi-metric (\hat{q})

when estimating from kNN and MNN-based estimators. The corresponding optimal number of neighbors k_{opt} is obtained by minimizing a global cross-validation criterion as in the simulated example before. Keep in mind that with a functional linear model, i.e., a model which does not involve any nonparametric component, the MSEP is 7.17. This value can serve as a benchmark for analyzing the results in Table 1. Table 1 indicates that the most suitable proximity measure is the semi-metric based on the second order derivatives:

$$d_2(X_i, X_j) = \sqrt{\int [X_i^{(2)}(t) - X_j^{(2)}(t)]^2 dt}.$$

The second derivatives of the absorbance curves are

kNN MSE=6.748



MNN MSE= 3.429

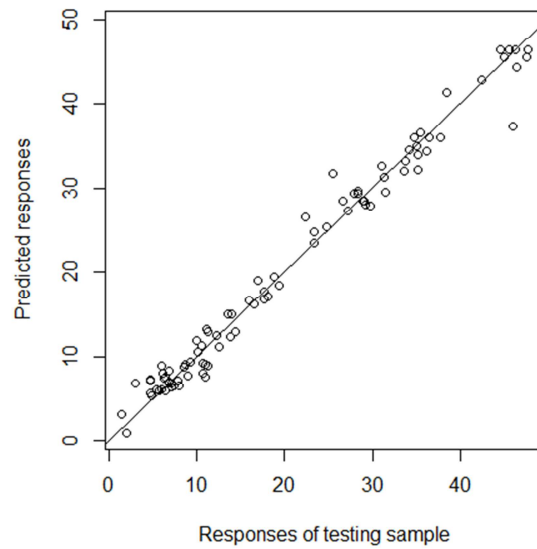


Figure 8. Knn method vs. mnn method for predicting fat content from spectrometric curves using the semi-metric based on the second derivatives.

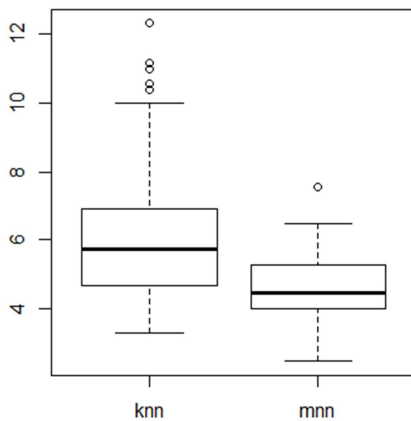


Figure 9. Boxplot of 100 values for the MSEP.

We only use a particular partition in the first attempt. To attenuate the effect of the considered partition, 100 partitions are generated at random. The semi-metric based on the second order derivatives has the best results and thus we use it as the proximity measure here. From each training sample and for both kNN and MNN methods, the tuning parameter k is

displayed in the right panel of Figure 7. Figure 8 displays the predicted values obtained by the two methods (kNN and MNN) on the testing sample with the semi-metric $d_2(\cdot, \cdot)$. The MNN method shows the better predictive power with a MSEP of 3.4294 (against 6.7480 for the kNN method).

Table 1. The selected semi-metric (\hat{q}) and the values of MSEP for kNN and MNN method.

\hat{q}	kNN	MNN
1	16.3255	11.4630
2	6.7480	3.4294
3	14.3117	14.4158
4	11.2274	7.3261

chosen as before and then Y_i ($i \in S_{test}$) is predicted. 100 values for the MSEP are obtained for each method and displayed as the boxplots in Figure 9. It clearly shows a better predictive power for the MNN method.

5. Conclusion

In this paper, we propose the mutual nearest neighbors (MNN) method in nonparametric functional regression. The center idea of MNN is that it identifies the mutual nearest neighbors and utilizes them to predict the response of the curve we concern. This strengthens the credibility of the final prediction result because the pseudo nearest neighbors are not taken into account during the prediction process. In addition, the L_2 -metric, a strict generalization of the Euclidean metric to functions, is not a good proximity measure sometimes. A solution is to allow a choice inside a large family of semi-metrics which will be built according to each specified statistical problem and dataset. A series of comparative experimental analyses show that MNN method can achieve better performance than kNN method with proper proximity measures.

References

- [1] Ramsay J O, Silverman B W. Functional Data Analysis [M]. Springer New York, 1997.
- [2] Ferraty F, Vieu P. Nonparametric functional data analysis: theory and practice [M]. Springer New York, 2006.
- [3] Goia A, Vieu P. An introduction to recent advances in high/infinite dimensional statistics [J]. Journal of Multivariate Analysis, 2016, 146 (2):1-6.
- [4] Wang J L, Chiou J M, Mueller H G. Review of Functional Data Analysis [J]. Statistics, 2015.
- [5] Morris J S. Functional regression [J]. Annual Review of Statistics and Its Application, 2015, 2: 321-359.
- [6] Reiss P T, Goldsmith J, Shang H L, et al. Methods for Scalar - on-Function Regression [J]. International Statistical Review, 2017, 85 (2): 228-249.
- [7] Royall R M. A class of non-parametric estimates of a smooth regression function [D]. Department of Statistics, Stanford University, 1966.
- [8] Stone C J. Consistent nonparametric regression [J]. The annals of statistics, 1977: 595-620.
- [9] Györfi L, Kohler M, Krzyzak A, et al. A distribution-free theory of nonparametric regression [M]. Springer Science & Business Media, 2006.
- [10] Laloë T. A k-nearest neighbor approach for functional regression [J]. Statistics & probability letters, 2008, 78 (10): 1189-1193.
- [11] Burba F, Ferraty F, Vieu P. k-Nearest Neighbour method in functional nonparametric regression [J]. Journal of Nonparametric Statistics, 2009, 21 (4): 453-469.
- [12] Gowda K C, Krishna G. Agglomerative clustering using the concept of mutual nearest neighbourhood [J]. Pattern recognition, 1978, 10 (2): 105-112.
- [13] Liu H, Zhang S, Zhao J, et al. A new classification algorithm using mutual nearest neighbors [C]. Grid and Cooperative Computing (GCC), 2010 9th International Conference on. IEEE, 2010: 52-57.
- [14] Guyader A, Hengartner N. On the mutual nearest neighbors estimate in regression [J]. The Journal of Machine Learning Research, 2013, 14 (1): 2361-2376.
- [15] Geenens G. Curse of dimensionality and related issues in nonparametric functional regression [J]. Statistics Surveys, 2011, 5: 30-43.