Consistence Condition of Kernel Selection In Regular Linear Kernel Regression And Its Application In Covid-19 High-risk Areas Exploration

WangChao (WangChaoAnHui@126.com)
LiBo
LiuHuan
PengPai

Research Article

Keywords: Linear Kernel, Covid-19 Source Exploration, Kernel Selection, Superimposed Fields

Posted Date: March 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1430675/v1

Consistence Condition of Kernel Selection In Regular Linear Kernel Regression And Its Application In Covid-19 High-risk Areas Exploration

Wang Chao[1*], Li Bo[1†], Liu Huan[1†] and Peng Pai[1†]

[1*]Department of Mathematic and Statistics, Central China Normal University, Luoyu, Wuhan, 430071, Hubei, China.

*Corresponding author(s). E-mail(s): WangChaoAnHui@126.com;
Contributing authors: haoyoulibo@163.com; 1083236002@qq.com; 469555176@qq.com;
†These authors contributed equally to this work.

Abstract

With the long-term outbreak of the Covid-19 around the world, identifying high-risk areas is becoming a new research boom. In this paper, we propose a novel regression method namely Regular Linear Kernel Regression(RLKR) for Covid-19 high-risk areas Exploration. We explain in detail how the canonical linear kernel regression method is linked to the identification of high-risk areas for Covid-19. Further more, The consistence condition of Kernel Selection, which is closely related to the identification of high-risk areas, is given with two mild assumptions. Finally, the RLKR method was verified by simulation experiments and applied for Covid-19 high-risk area Exploration.

Keywords: Linear Kernel; Covid-19 Source Exploration; Kernel Selection; Superimposed Fields

1 Introduction

Covid-19 was recognized in December 2019[1]. Since then, the Covid-19 outbreak has posed critical challenges for the public health, research, and medical communities. Currently, it was recognized that human to human transmission played a major role in the subsequent outbreak [2]. Based on this Knowledge, in the data collection stage, we try to extract data related to the flow of people as much as possible. At present, the research on the Covid-19 mainly focuses on the treatment of the epidemic[3, 4], the transmission of the epidemic[5, 6] and the prevention of the epidemic[7, 8]. In F. Ndaïrou 2020, the concept of super-spreaders, who contribute disproportionately to a much larger number of cases in the ongoing Covid-19 pandemic, was proposed for modeling the Covid-19 transmission. We focus our attention on high-risk areas exploration from "sky" perspective. Which means that, we ignore different regional political factors and medical level factors. All the factors we can known come from aerial monitoring. We imagine ourselves as a space traveler. Before landing on the earth, determining which areas are high-risk areas and which areas are low-risk areas through detection in the sky is actually our task. In order to make a decision, we had to count the number of outbreaks in each region, as well as all the data related to the spread of the outbreak, e.g. population, population density per square kilometer, population transmission, people gathering frequency, distances between regions, etc. Given this information of different areas, we cannot decide the risk level of one area, just by the outbreaks. Because, for several adjacent areas, the surrounding areas are likely to be only the victim, and only one area is the source of the outbreak. One the other hand, for areas of high volume of population transmission and people gathering frequency, its low-risk label may only be temporary. Along the continuous outbreak of Covid-19, those areas are much like to become high-risk areas.

The inspiration of Regular Linear Kernel Regression, which is used for high-risk area detection of Covid-19 in this paper, comes from a variable selection method, called Lasso[9], which is firstly proposed for solving S-sparse linear regression model in high-dimensional statistics. Actually the proof framework of Kernel Selection Consistence origins from the variable selection consistence of Lasso[10], but a slightly modification for suitable for Kernel Selection. To gain an intuition of the link between Kernel Selection and high-risk areas exploration, let us see inside of Kernel Regression. Denote $K_\alpha(x) = K(\alpha, x)$ be a measure of the distance between α and x, where the Gaussian RBF(Radial Basis Function) Kernel and Linear Inner Product Kernel are most widely used. In this paper, the kernel selection consistence condition of Regular Linear Kernel Regression was given. For a given group of kernels, $K_\alpha(\cdot), \alpha \in \bar{\alpha}$, where $\bar{\alpha}$ stands for an index set, Kernel Regression model has the form,

$$y = \sum_{\alpha \in \bar{\alpha}} \theta_\alpha K_\alpha(x) + w, \tag{1}$$

where, y stands for outbreaks and w stands for observation noise. From the model, y is actually a linear combination of the characteristic Kernels at the point of x. We see a Kernel $K_\alpha(x)$ as a spread function of a high-risk area characterized by α. As stated, $K_\alpha(x)$ is a measure of the distance between α and x. For a long distance area x from high-risk area α, the influence from α, source of transmission, shall be small, which is coincident with the property of $K_\alpha(x)$. Upper to now, we explain the link between the spreading mode of Covid-19 and the Kernel Regression, that's high-risk areas were seen as emission sources characterized by $K_\alpha(\cdot)$ in the model and the outbreaks in one area is actually the combination of the influences of those emission sources. Still, there are problems that must be check carefully. First, the number of high-risk areas can be unlimited? For a given design matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$, and a candidate set $Card(\bar{\alpha}) = m$, the number of high-risk areas s must be less than or equal to $\min\{n, d, m\}$, other wise the true high-risk areas will not be check out entirely. Second, for a given estimation of $\{\hat{\theta}\}$ with a support set of $\hat{S}$, under which conditions the support set is consistent with the true support set S of true θ? This question is crucial, as it means whether we have got the true high-risk areas or Covid-19 sources. As a non-zero θ_α means that the candidate area of $K_\alpha(\cdot)$ influences other areas.

Notice that, Interpolation learning[11, 12] has recently attracted growing attention in the machine learning community. Linear Kernel Regression model has been around for decades, whoes continued popularity stems from its excellent properties, well generalization along with good interpolation, in high-dimensional statistics. Due to the derivability, Kernel Ridge Regression with l_2-norm penalization is very popular in the High-dimensional area, including the converge rate[13], generalization risk[14] and variable selection[15]. While the Consistence condition of Kernel Regression with l_1-norm penalization is the novelty of this paper, that's the topic of Kernel selection consistency under l_1-norm regularization is first proposed in this paper. We compliment the consistence condition for Kernel Regression with l_1-norm penalization, along with the generalization risk. Due to the interpolation of Kernel regression[16], the connection between Kernel selection with high-risk areas detection of Covid-19 is our innovation. Around those question, we structured our paper as follows. In the introduction, we briefly introduce the background of Covid-19 transmission and give the inspiration of high-risk areas exploration. In the second section, some remarks were introduced and Linear Kernel Regression Model was propose for the spreading mode of Covid-19 on high-risk areas influence view. As we have no hope for estimating an unique solution of $\hat{\theta}$ with the optimal least square method. Regular technique was suggested for estimating an unique solution. In the third section, the conditions of Kernel Selection Consistence were given. Finally the simulation experiment was conduct for supplements of consistence checking. Then the Kernel Selection Method was applied for Covid-19 high-risk areas detection with outbreaks of Covid-19 up to year 2021 along with some other features that influence the transmission of Covid-19.

2 Model Construction

Suppose we are given relevant features in observation regions, $x_i = (x_{i1}, x_{i2}, \cdots, x_{id})^T$ and corresponding response values y_i, for $i = 1, \cdots, n$, which stands for the most concerned variable, e.g. the number of epidemic outbreaks. A significant research is to find possible emission sources or high-risk areas from many candidate areas $\alpha_j = (\alpha_{j1}, \cdots, \alpha_{jd})^T$, for $j = 1, \cdots, m$. The true high-risk areas α_S is supposed to be contained in those candidate areas $\bar{\alpha} = [\alpha_1^T, \cdots, \alpha_m^T]^T$ indexed by S and $S \subseteq [m]$ stands for the true support of high-risk areas. Here $[m]$ is a simplified notation of $\{1, \cdots, m\}$.

Intuitively, the response variable y at observation point x is the linear superimposed effects of each real high-risk areas, that's

$$y = \sum_{j \in S} \theta_j < x, \alpha_j > + w, \tag{2}$$

where $S \subseteq \{1, \cdots, m\}$ stands for the supports of truth emission sources, and w stands for observation noise. $< \cdot, \cdot >$ stands for inner product. Since S is unknown in practice, the estimation problem of S will be more meaningful, except for the estimation of $\hat{\theta}_j, j \in S$. Actually if the support S of the truth emission sources was given, we can estimate $\hat{\theta}$ very well. With the true S unknown, there is no hope for obtaining an unique solution of θ_j, when $m \geq \min\{n, d\}$, from Lemma 1 in Appendix. Without uniqueness, there is no way to talk about the Kernel Selection Consistence. Inspired by Lasso of l_1 norm regularization, consider the following optimization,

$$\hat{\theta} \leftarrow \arg\min \sum_{i=1}^{n} \frac{1}{2n} (y_i - \sum_{j \in [m]} \theta_j < x_i, \alpha_j >)^2 + \lambda_n \|\theta\|_1. \tag{3}$$

Here, $[m] = \{1, \cdots, m\}$ is a simple notation and λ_n is a given parameter. Denote $\mathbf{X} = [x_1, \cdots, x_n]^T \in \mathbb{R}^{n \times d}$, $\alpha = [\alpha_1, \cdots, \alpha_m]^T \in \mathbb{R}^{m \times d}$, $\theta = [\theta_1, \cdots, \theta_m]^T \in \mathbb{R}^m$ and $\mathbf{Y} = [y_1, \cdots, y_n]^T \in \mathbb{R}^n$. Therefore a matrix form alternative of equation (3), namely Regular Linear Kernel Regression, is given by,

$$\hat{\theta} \leftarrow \arg\min \frac{1}{2n} \|\mathbf{Y} - \mathbf{X}\alpha^T\theta\|_2^2 + \lambda_n \|\theta\|_1. \tag{4}$$

For the not differentiable of l_1-norm, due to its sharp point at the origin, there's no hope for a close form solution. Since this is a convex optimization problem, many optimization methods can be used to solve the problem, e.g. Gradient descent or Least Angel Regression.

So the method of estimating $\hat{\theta}$ is not the focus of this paper. The explanation of the estimation result of $\hat{\theta}$ and the condition of consistence between the estimation and the truth θ is our focus. Here we give the explanation of $\hat{\theta}$. From the model (2), for a zero $\theta_j = 0$, which means y is irrelevant with α_j, the location of α_j shall not be the emission source. For a non-zero θ_j, which

means y is relevant with α_j, the location of α_j shall be the emission source. For a positive θ_j, the corresponding location of α_j shall be see as pollution source in the pathogen exploration case. And a negative θ_j, the corresponding location of α_j shall be see as purification source.

3 Condition Of Consistence

Consider an S-sparse linear kernel regression model and the true support of $\theta^\star$ is S. Given an estimation $\hat{\theta}$ with support $\hat{S}$. The question is by which condition the estimation $\hat{S}$ is consistent with S? We begin by assuming $\mathbf{X}$ be deterministic design matrices. Let S stands for the supports of truth $\theta^\star$ and α_S stands for the subset of $\alpha = [\alpha_1, \cdots, \alpha_m]^T$ with indices of S. First the truth emission source α_S cannot be linearly correlated, in order to ensure that the model is identifiable, even if the support set S were known a priori. Second the fake source α_{S^C} shall be irrelevant with the truth emission source α_S, which makes it possible to exclude irrelevant areas. In general, we cannot expect this orthogonality to hold, but a type of approximate orthogonality. In particular, consider the following conditions:

(A1) *Lower eigenvalue* : The smallest eigenvalue of the sample covariance submatrix indexed by S is bounded below:

$$\gamma_{min}\left(\frac{\alpha_S \mathbf{X}^T \mathbf{X} \alpha_S^T}{n}\right) \geq c_{min} > 0. \tag{5}$$

(A2) *Mutual incoherence* : There exists some $p \in [0, 1)$ such that

$$\max_{j \in S^C} \|(\alpha_S \mathbf{X}^T \mathbf{X} \alpha_S^T)^{-1} \alpha_S \mathbf{X}^T \mathbf{X} \alpha_j\|_1 < p. \tag{6}$$

With this set-up, the following results follow applied to the lagrange form Linear Kernel Lasso (4) when applied to an instance of the linear kernel model (2) such that the true parameter $\theta^\star$ is supported on a subset S with cardinality s. In order to state the result, we introduce the convenient shorthand $\Pi_{S^\perp}(\mathbf{X}\alpha^T) = \mathbf{I}_n - \mathbf{X}\alpha_S^T(\alpha_S \mathbf{X}^T \mathbf{X} \alpha_S^T)^{-1}\alpha_S \mathbf{X}^T$, a type of orthogonal projection matrix.

Theorem 1 *Consider an S-sparse linear kernel regression model for which the design matrix satisfies conditions (5) and (6). Then for any choice of regularization parameter such that*

$$\lambda_n \geq \frac{2}{1-p} \left\| \alpha_{S^C} X^T \Pi_{S^\perp(\boldsymbol{X}\alpha^T)} \frac{w}{n} \right\|_\infty, \tag{7}$$

the Lagrangian Kernel Lasso (3) has the following properties:
(a) Uniqueness: There is a unique optimal solution $\hat{\theta}$.
(b) No false inclusion: The solution has its support set $\hat{S}$ contained within the true

support set S.
(c) l_∞*-bounds: The error* $\hat{\theta} - \theta^\star$ *satisfies*

$$\|\hat{\theta}_S - \theta^\star_S\|_\infty \leq \underbrace{\left\| \left(\frac{\alpha_S \boldsymbol{X}^T \boldsymbol{X} \alpha_S^T}{n} \right)^{-1} \alpha_S \boldsymbol{X}^T \frac{w}{n} \right\|_\infty + \left\| \left(\frac{\alpha_S \boldsymbol{X}^T \boldsymbol{X} \alpha_S^T}{n} \right)^{-1} \right\|_\infty \lambda_n}_{B(\lambda_n;\ \boldsymbol{X}\alpha^T)} . \quad (8)$$

(d) No false exclusion: The solution's support set $\hat{S}$ *includes all indices* $i \in S$ *such that* $| \theta_i^\star | > B(\lambda_n;\ \boldsymbol{X}\alpha^T)$*, and hence is sample selection consistent if* $\min_{i \in S} | \theta_i^\star | > B(\lambda_n;\ \boldsymbol{X}\alpha^T)$.

The prove of this result utilizes a technique called a primal-dual witness method, which constructs a solution and then the solution was shown be optimal and unique, see Appendix. Here let us try to interpret its main claims. First the uniqueness claim in part (a) is not trivial. Although the Lagrange objective is convex, it can never be strictly convex when $m > \min\{d, n\}$. Based on the uniqueness claim, we can talk unambiguously about the support of the Kernel Lasso estimate $\hat{\theta}$. Part (b) guarantees that the kernel lasso does not falsely include linear kernels that are not in the support of $\theta^\star$, or equivalently that $\hat{\theta}_{S^C} = 0$. Part (d) is a consequence of the l_∞-norm bound from part (c): as long as the minimum value of $| \theta_i^\star |$ over indices $i \in S$ is not too small, then the Kernel Lasso is kernel selection consistent in the full sense.

Further more, if we make specific assumptions about the noise vector w, more concrete results can be obtained. Before giving the results, we give the definition of the Sub-Gaussian variable.

Definition 1 A random variable w with mean $\mu = \mathbb{E}[w]$ is sub-Gaussian if there is a positive number σ such that

$$\mathbb{E}[e^{\lambda(w-\mu)}] \leq e^{\sigma^2 \lambda/2} \qquad for \quad all \quad \lambda \in \mathbb{R}.$$

With this in mind, we given a third assumption, that's

(A3) The observation noise $w_i, i \in \{1, \cdots, n\}$ is i.i.d. σ-sub-Gaussian variables.

Before given the concrete corollary about Linear Kernel Regression, a Lemma about lower bounds for sub-Gaussian variables is needed.

Lemma 2 Let $\{Z_i\}_{i=1}^n$ be a sequence of zero-mean random variables, each sub-Gaussian with parameter σ. (No independence assumptions are needed.) Consider the lower bound of $Z = \max_{i=1,\cdots,n} \mathsf{X\_i}$, we have

$$\mathbb{P}\left[Z \geq t\right] \leq 2ne^{-\frac{t^2}{2\sigma^2}}, \quad (9)$$

valid for all $n \geq 2$.

Corollary 3 Consider the S-sparse linear kernel regression model based on a noise vector w that assumption (A3) holds, and a deterministic design matrix $\mathbf{X}$ that satisfies assumptions (A3) and (A4). And the character Kernel satisfies the C-Column normalization condition: $\max_{k=1,\cdots,m} \|\mathbf{X}\alpha_k\|_2/\sqrt{n} \leq C$. Suppose that we solve the Regular Linear Kernel Regression (3) with regularization parameter

$$\lambda_n = \frac{2C\sigma}{1-p}\left\{\sqrt{\frac{2\log(m-s)}{n}} + \delta\right\} \tag{10}$$

for some $\delta > 0$. Then the optimal solution $\hat{\theta}$ is unique with its support contained within S, and satisfies the l_∞-error bound

$$\|\hat{\theta}_S - \theta^\star_S\|_\infty \leq \frac{\sigma}{\sqrt{c_{min}}}\left\{\sqrt{\frac{2\log s}{n}} + \delta\right\} + \left\|\left(\frac{\alpha_S \mathbf{X}^T\mathbf{X}\alpha_S^T}{n}\right)^{-1}\right\|_\infty \lambda_n, \tag{11}$$

with probability at least $1 - 4e^{-\frac{n\delta^2}{2}}$.

Proof First, let us show that the given choice of λ_n (10) of the regularization parameter satisfies the bound (7) with high probability. It suffices to bound the maximum absolute value of the random variables

$$Z_k := \alpha_k^T\mathbf{X}^T \underbrace{[\mathbf{I}_n - \mathbf{X}\alpha_S^T(\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T)^{-1}\alpha_S\mathbf{X}^T]}_{\Pi_{S^\perp}(\mathbf{X})}\left(\frac{w}{n}\right) \qquad for \quad k \in S^c.$$

Since $\Pi_{S^c}(\mathbf{X})$ is an orthogonal projection matrix, we have

$$\|\Pi_{S^\perp}\mathbf{X}\alpha_k\|_2 \leq \|\mathbf{X}\alpha_k\| \leq C\sqrt{n},$$

where the last inequality follows from the C-Column normalization condition. Therefore, each variable Z_j is sub-Gaussian with parameter at most $C^2\sigma^2/n$. By Lemma 2, it holds that,

$$\mathbb{P}\left[\max_{j\in S^c} \mathrm{Z\_j} \geq t\right] \leq 2(m-s)e^{-\frac{nt^2}{2C^2\sigma^2}}.$$

Transforming this to the l_∞-norm form, we have

$$\mathbb{P}\left[\frac{2}{1-p}\left\|\alpha_{S^C}X^T\Pi_{S^\perp(\mathbf{X}\alpha^T)}\frac{w}{n}\right\|_\infty \geq \lambda_n\right] \leq 2(m-s)e^{-\frac{n(1-p)^2\lambda_n^2}{8C^2\sigma^2}}.$$

Substitute the choice of λ_n (10), we have

$$\lambda_n \geq \frac{2}{1-p}\left\|\alpha_{S^C}X^T\Pi_{S^\perp(\mathbf{X}\alpha^T)}\frac{w}{n}\right\|_\infty,$$

with probability at least $1 - 2e^{-\frac{n\delta^2}{2}}$.

It remains to simplify the l_∞-bound (8). Consider the random variables $\tilde{Z}_i := e_i^T\left(\frac{1}{n}\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T\right)^{-1}\mathbf{X}\alpha_S^T w/n$, Since the elements of the vector w are i.i.d. σ-sub-Gaussian, the variable $\tilde{Z}_i$ is zero-mean and sub-Gaussian with parameter at most

$$\frac{\sigma^2}{n}\left\|\left(\frac{1}{n}\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T\right)^{-1}\right\|_2 \leq \frac{\sigma^2}{c_{\min}n},$$

where we have used the eigenvalue condition (5). Consequently, for any $\delta > 0$, we have

$$\mathbb{P}\left[\max_{i=1,\cdots,s} \mid \tilde{Z}_i \mid > \frac{\sigma}{\sqrt{c_{\min}}}\left\{\sqrt{\frac{2\log s}{n}} + \delta\right\}\right] \leq 2e^{-\frac{n\delta^2}{2}}.$$

Then the claims follows from $(1 - 2e^{-\frac{n\delta^2}{2}})^2 \geq 1 - 4e^{-\frac{n\delta^2}{2}}$. □

Upper to here, we have given the Kernel selection consistence conditions. By careful analysis of the original optimization problem (3), we find that the values of $\{\hat{\theta}_k\}_{k=1}^m$ gradually diminish, as the penalize parameter λ_n is large enough. From condition (7), if the given parameter λ_n is too small, (7) condition is violated, we have no hope for excluding all fake Kernels. From part (d) of Theorem 1, the $\hat{\theta}_i, i \in S$, which cannot be zero, may be miscalculated as zero. In real world data detection, the choice of λ shall be carefully checked. Besides, Cross-validation methods are usually utilized for the decision of λ_n, or other improved methods [17].

4 Prediction Error Or Risk Assessment

Another interesting twist is to do a risk assessment for all areas. Consider the Linear Kernel Regression Model (2). Given an estimation of $\hat{\theta}$, the prediction of outbreaks $\hat{y}$ is the superposition summation of the Kernels at the position of x, that's

$$\hat{y} = \sum_{k\in[m]} \hat{\theta}_k < \alpha_k, x > .$$

As this prediction $\hat{y}$ is the linear combination of the Kernel functions, radiation effects of the true high-risk areas, at the point x, not just the number of the local outbreaks. We see $\hat{y}$ as the risk assessment of the area x. Given a S-sparse Linear Kernel Regression model, we devoted ourselves to bound on the mean-squared prediction error

$$\frac{\|\mathbf{X}\alpha^T(\hat{\theta} - \theta^\star)\|_2^2}{n} = \frac{1}{n}\sum_{i=1}^n \left[\sum_{k=1}^m < \alpha_k, x_i > (\hat{\theta}_k - \theta_k^\star)\right]^2. \tag{12}$$

Before giving the results, some more constraints about the design matrix, namely restricted eigenvalue (RE) condition shall be advertised here.

Definition 2 The matrix $\mathbf{X}$ satisfies the restricted eigenvalue (RE) condition over S with parameters (κ, γ) if

$$\frac{1}{n}\|\mathbf{X}\Delta\|_2^2 \geq \kappa\|\Delta\|_2^2 \qquad for \quad all \quad \Delta \in \mathbb{C}_\gamma(S).$$

Here the subset $\mathbb{C}_\gamma(S) = \{\Delta \in \mathbb{R}^d \mid \|\Delta_{S^c}\|_1 \leq \gamma\|\Delta_S\|_1\}$, is a collection of vectors whose l_1-norm off the support is dominated by the l_1-norm on the support.

The restricted eigenvalue (RE) condition constructs a link between the recovering error $\|\hat{\theta} - \theta^\star\|_2$ and the prediction error $\|\mathbf{X}\alpha^T(\hat{\theta} - \theta^\star)\|_2$. With this set-ups, we give the bounds on prediction errors as follows.

Theorem 4 *(Prediction error bounds) Consider the Regular Linear Kernel Regression Optimization Problem (3) with a strictly positive regularization parameter* $\lambda_n \geq 2\|\frac{\alpha \boldsymbol{X}^T w}{n}\|_\infty$.
(a) Any optimal solution $\hat{\theta}$ *satisfies the bound*

$$\frac{\|\boldsymbol{X}\alpha^T(\hat{\theta} - \theta^\star)\|_2^2}{n} \leq 12\|\theta^\star\|_1\lambda_n.$$

(b) If $\theta^\star$ *is supported on a subset* S *of cardinality* s*, and the design matrix* $\boldsymbol{X}\alpha^T$ *satisfies the* $(\kappa;\ 3)$*-RE condition over* S*, then the optimal solution satisfies the bound*

$$\frac{\|\boldsymbol{X}\alpha^T(\hat{\theta} - \theta^\star)\|_2^2}{n} \leq \frac{9}{\kappa}s\lambda_n^2.$$

The proof of this prediction bounds can be seen in Appendix. Before that, let's see what we are informed from the bounds. The two upper bounds are all related to λ_n. In order to reduce the forecast errors, a critical step is to reduce the regularization parameter λ_n.

5 Simulation Studies

Consider a S-sparse Linear Kernel Regression model (2). Suppose each row x_i of the design matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ is independently draw from a Gaussian distribution $N(0, \mathbf{I}_d)$. And each α_j of the candidates kernels $\alpha \in \mathbb{R}^{m\times d}$ is also i.i.d of Gaussian distribution $N(0, \mathbf{I}_d)$. The true support S is a subset of $\{1, \cdots, m\}$ uniformly distributed in $[m]$. The real parameter of $\theta_j^\star, j \in S$ is uniformly distributed on $[-1, -0.1] \cup [0.1, 1]$, and $\theta_j^\star = 0$ for $j \in S^c$. Let the entries of the noise vector w follow i.i.d of a Gaussian distribution of $N(0, 0.01)$. Then the response variable Y is calculated by

$$Y = \mathbf{X}\alpha^T\theta + w.$$

Applied by the regular method of l_1-norm, we are interested in the false-inclusion rate and consistency. False-inclusion rate is defined as,

$$False - inclusion = \frac{\hat{S}\backslash S}{\hat{S}}, \tag{13}$$

where $\hat{S}$ stands for the support of estimation $\hat{\theta}$. And the consistency between $\hat{S}$ and S, is given by

$$Consistency = \frac{\hat{S} \cap S}{\hat{S} \cup S}. \tag{14}$$

Since the choice of λ_n is critical for consistence. We checked the false-inclusion

Table 1 The False-inclusion rate and consistency of given levels of λ_n, where $n = m = 1000$, $d = 500$.

s	λ_n	F	Con	s	λ_n	F	Con
$s = 5$	$0 \sim 0.03$	0.992	0.007	$s = 15$	$0 \sim 0.05$	0.97	0.029
	$0.03 \sim 55.5$	0	1		$0.05 \sim 68.7$	0	1
	$55.5 \sim 95.9$	0	0.8		$68.7 \sim 72.7$	0	0.933
	$95.9 \sim 242$	0	0.6		$72.7 \sim 103$	0	0.867
$s = 10$	$0 \sim 0.10$	0.988	0.011	$s = 20$	$0 \sim 0.05$	0.964	0.035
	$0.10 \sim 103$	0	1		$0.05 \sim 82.8$	0	1
	$103 \sim 254$	0	0.9		$82.8 \sim 105.5$	0	0.95
	$254 \sim 263.6$	0	0.8		$105.5 \sim 128.3$	0	0.9

Table 2 The average False-inclusion rate and consistency of dimensions of $d \in [40, 100, 200, 300, 500]$. Here $n = 1000$, $s = 20$ and the average is calculated over 100 replicates.

m	d	F	Con	m	d	F	Con
$m = 40$	$d = 40$	0.205	0.772	$m = 300$	$d = 40$	0.603	0.234
	$d = 100$	0.043	0.956		$d = 100$	0.330	0.593
	$d = 200$	0.004	0.995		$d = 200$	0.074	0.923
	$d = 300$	0.001	0.998		$d = 300$	0.014	0.984
	$d = 500$	0	1		$d = 500$	0.001	0.999
$m = 100$	$d = 40$	0.462	0.430	$m = 1000$	$d = 40$	0.731	0.118
	$d = 100$	0.158	0.830		$d = 100$	0.475	0.366
	$d = 200$	0.019	0.980		$d = 200$	0.169	0.800
	$d = 300$	0.002	0.997		$d = 300$	0.049	0.948
	$d = 500$	0	1		$d = 500$	0.002	0.997

(F) and consistency (Con) under given levels of λ_n in Table 1. From Table 1, we known that the best result is $F = 0$, which means no false inclusion and $Con = 1$, which means the estimation support $\hat{S}$ is identical with the true kernel support S. For λ_n that's not too small, the no false inclusion holds. And the consistence grows up and down, as λ_n increases. The consistence holds for a long time as we increase the value of λ_n, which means that for a long period, the support of $\hat{\theta}$ makes no changes. And this information will be useful for empirical selection of the regularization parameter λ_n.

Further, we find that the sample dimension has an important relationship with the consistency. We calculated the average consistency under different dimensions of $d \in [40, 100, 200, 300, 500]$. The average is the result of 100 replicates, where the λ_n is chosen according to maximizing consistence. From the results in Table 2, it's easy to find that as the dimension increases, the average consistency increases steadily. As the number of candidate Kernels increase, picking out exactly the true supports becomes more difficult. From here, we get the revelation that increasing the dimension of the data is of great significance to correctly pick out the Kernels. And reducing the number of candidate kernels is also useful for reducing $false-inclusion$ rates and improving $consistency$.

Finally, We care about the changing law of the prediction error, by adding tests check. We draw the test data-set by choosing a test design matrix $\mathbf{X}^{test} \in \mathbb{R}^{T \times d}$, with each row x_i^{test} follows a Gaussian distribution $N(0, \mathbf{I}_d)$. And the

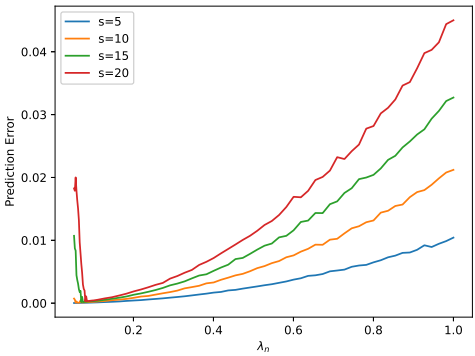

Fig. 1 The Prediction error of Regular Linear Kernel Regression.

test response variable Y^{test} is calculated by

$$Y^{test} = \mathbf{X}^{test}\alpha^T\theta + w.$$

The prediction error is calculated by equation (12). The curves of prediction error along λ_n were plotted in Figure 1 under given levels of $s \in [5, 10, 15, 20]$, which stands for the number of true Kernels. Here $n = 1000, m = 1000, d = 500, T = 500$, where T stands for the size of test data-set. From the results of the Figure 1, the prediction error grows as λ_n goes beyond some fixed point. And for the different levels of s, the prediction error grows proportional to s, which is coincident with the result from part (b) of Theorem 4.

6 High-risk Areas Selection and Risk Assessment of Covid-19

In this section, We try to explore high-risk areas of Covid-19 with the outbreaks of Covid-19, provided by https://fy.onesight.com/. This data includes the number of confirmed cases in each country since the outbreak of Covid-19. As the data collection deadline is September 2021, we had to collect the relevant covariates of each country up to that point in time. First, we collected the coordinates of the capital of each country, from the latitude β_t and longitude β_g, by

$$\begin{cases} x = \cos(\beta_t)\cos(\beta_g), \\ y = \cos(\beta_t)\sin(\beta_g), \\ z = \sin(\beta_t). \end{cases}$$

As mentioned earlier, the Covid-19 now spreads mainly in the form of person-to-person spread. So in the second part of date collection, the variables of population transmission were collected, including population, population density per square kilometer, population birth rate and Gross Domestic Product (GDP), which concerns the activity of each country.

Table 3 The high-risk areas and purification areas selected by Regular Linear Kernel Regression.

λ_n	$high-risk\ area$	$purification\ area$
$\lambda_n = 2$	United States, India, Malaysia, Indonesia	Brazil, Russia, Mexico, Peru, China
$\lambda_n = 1$	United States, India, Malaysia, Indonesia,	Brazil, United Kingdom, Russia, Mexico, Peru, Iraq, Czechia, Chile, Israel, Portugal, Jordan, United Arab Emirates, Oman, Bahrain, Equatorial Guinea, Saudi Arabia, China, Italy, Monaco.
$\lambda_n = 0.5$	United States, India, Malaysia, Indonesia, Turkey, Japan	Brazil, United Kingdom, Russia, Mexico, Peru, Iraq, Czechia, Chile, Israel, Portugal, Jordan, United Arab Emirates, Oman, Bahrain, Equatorial Guinea, Saudi Arabia, China, Italy, Monaco. Poland, Ukraine, Netherlands, Czechia, Libya, Egypt, Zambia, Democratic Republic of Congo

And in the third part of the data collection, the history of death toll of Covid-19 by weeks is also collected from https://www.kaggle.com/tarunkr/covid-19-case-study-analysis-viz-comparisons from January 2020 to September 2021. In summary, we collected outbreaks and associated covariates of 114 dimension for 189 countries.

The first concern is the selection of high-risk areas. We makes all 189 countries or regions be candidates of Kernels. Before the regular Linear Kernel Regression model being implemented, a normalize scale of data set is employed, as follows

$$x_{scale} = \frac{x - \bar{x}}{\sqrt{S^2(x)}}.$$

Here, $\bar{x}$ stands for sample mean of x and $S^2(x)$ stands for sample covariates, $S^2(x) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. Given several levels of λ_n, the Kernel areas selected were given in Table 3. In the table, high-risk areas are countries , whose coefficients are positive. Purification areas are countries, whose coefficients are negative.

Another meaningful task is to carry out a risk assessment of all candidate areas. Given $\lambda_n = 0.5$, we made a risk assessment for all 189 countries or areas. The risk is calculated by

$$Y_{risk} = \sum_{k=1}^{m} \hat{\theta}_k < x, \alpha_k > .$$

As all variables of covariates were scaled with zero-mean and standard deviation $S^2(x) = 1$, including the number of outbreaks. So the risk Y_{risk} of each area takes values around 0. For high-risk areas, the response Y_{risk} is also large. In particularly, the assessment results are presented in Table 4, 5. From the risk assessment table, we see that Brazil, United Kingdom and Russia are of high risk in risk assessment program. While in purification areas selection, they

are viewed as purification areas. The reason lies in the those areas are close in "distance" with high-risk areas and actually they are victims. They are surrounded by the true Covid-19 emission sources. The relation between high-risk selection with risk assessment is that high-risk selection is finding the causes or emission sources and risk assessment is the presentation of spreading of Covid-19. For some areas that are judged to be of great risk assessment, they are likely to be innocent and may be purification areas. With Kernel selection of high-risk areas, Only then can we decide whether an area is a true source of transmission, not just by risk assessment.

7 Conclusion

In this paper, we considered the condition of Kernel selection consistence, in Regular Linear Kernel Regression Model with l_1-norm. And bounds on the prediction error of RLKR are also given with mild conditions. Then we applied the RLKR to high-risk selection and risk assessment for Covid-19, in a view point of area-to-area spreading mode. In the simulation of random design matrix, we explored the $False - inclusion$ and $Consistency$ under different levels of λ_n and the number m of candidate kernels. The prediction error bounds were also checked with a no-wise result. In real world data of Covid-19, the risk assessment of 189 number of countries or areas were given. We believe this will provide valid advices on traveling.

Acknowledgments. Our thanks go to Professor Bo Li for stimulating this research. Our research is supported by NSF of China (Grant No. 61877023) and the Fundamental Research Funds for the Central Universities (Grant No. CCNU19TD009).

Appendix A

Proof **(Theorem 1)** As the non-differentiable of l_1 norm, dual to the sharp point at 0. A tool named subgradient is needed. Given a convex function $f : \mathbb{R}^m \to \mathbb{R}$, we say $z \in \mathbb{R}^m$ is the subgradient of f at θ, denoted by $z \in \partial f(\theta)$, if we have

$$f(\theta + \Delta) \geq f(\theta) + < z, \Delta > \qquad for \quad all \quad \Delta \in \mathbb{R}^m.$$

For $f(\theta) = \|\theta\|_1$, it holds that $z \in \partial\|\theta\|_1$ if and only if $z_j = sign(\theta_j)$, for $j = 1, \cdots, m$. Here $sign(0)$ can be any number in the interval $[-1, 1]$. For Regular Linear Kernel Regression program (3), we say that a pair $(\hat{\theta}, \hat{z}) \in \mathbb{R}^m \times \mathbb{R}^m$ is primal-dual optimal if $\hat{\theta}$ is a minimizer and $z \in \partial\|\theta\|_1$. Any such pair must satisfy the zero-subgradient condition

$$\frac{1}{n}\alpha\mathbf{X}^T(\mathbf{X}\alpha^T\hat{\theta} - \mathbf{Y}) + \lambda_n\hat{z} = 0, \tag{15}$$

which is the analog of a zero-gradient condition in the non-differentiable setting. For the existence and uniqueness, we constructs a pair $(\hat{\theta}, \hat{z})$ satisfying the zero-subgradient condition (15), and such that $\hat{\theta}$ has the correct signed support. The constructive procedure namely Primal Dual Witness (PDW) is as follows

1. Set $\hat{z}_{S^c} = 0$.

2. Determine $(\hat{\theta}_S, \hat{z}_S) \in \mathbb{R}^s \times \mathbb{R}^s$ by solving the oracle subproblem

$$\hat{\theta}_S \leftarrow \arg\min_{\theta_S \in \mathbb{R}^s} \left\{ \underbrace{\frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\alpha_S^T\theta_S\|_2^2}_{f(\theta_S)} + \lambda_n \|\theta_S\|_1 \right\}. \tag{16}$$

   And choose $\hat{z}_S \in \partial\|\theta_S\|_1$ such that $\nabla f(\theta_S)\mid_{\theta_S=\hat{\theta}_S} + \lambda_n \hat{z}_S$.
3. Solve for $\hat{z}_{S^c} \in \mathbb{R}^{m-s}$ via the zero-subgradient equation (15), and check whether or not the strict dual feasibility condition $\|\hat{z}_{S^c}\|_\infty < 1$ holds.

We say the constructive procedure succeed if the strict dual feasibility condition holds in the last step of the program. From Lemma (5), we have the unique solution $(\hat{\theta}_S, 0)$ for the program (3). It remains to show the strict dual feasibility condition must be hold, that's $\hat{z}_{S^c} < 1$. Let's give a block matrix form of zero-subgradient condition first, that's

$$\frac{1}{n}\begin{bmatrix} \alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T & \alpha_S\mathbf{X}^T\mathbf{X}\alpha_{S^c}^T \\ \alpha_{S^c}\mathbf{X}^T\mathbf{X}\alpha_S^T & \alpha_{S^c}\mathbf{X}^T\mathbf{X}\alpha_{S^c}^T \end{bmatrix}\begin{bmatrix} \hat{\theta}_S - \theta_S^\star \\ 0 \end{bmatrix} - \frac{1}{n}\begin{bmatrix} \alpha_S\mathbf{X}^T w \\ \alpha_{S^c}\mathbf{X}^T w \end{bmatrix} + \lambda_n \begin{bmatrix} \hat{z}_S \\ \hat{z}_{S^c} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{17}$$

Solving this condition, we have

$$\hat{z}_{S^c} = -\frac{1}{\lambda_n n}\alpha_{S^c}\mathbf{X}^T\mathbf{X}\alpha_S^T(\hat{\theta}_S - \theta_S^\star) + \alpha_{S^c}\mathbf{X}^T\left(\frac{w}{\lambda_n n}\right).$$

Similarly, use the invertibility of $\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T$, we have

$$\hat{\theta}_S - \theta_S^\star = (\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T)^{-1}\alpha_S\mathbf{X}^T w - \lambda_n n(\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T)^{-1}\hat{z}_S. \tag{18}$$

Substituting this expression into the calculation of $\hat{z}_{S^c}$, we must have

$$\hat{z}_{S^c} = \underbrace{\alpha_{S^c}\mathbf{X}^T\mathbf{X}\alpha_S^T(\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T)^{-1}\hat{z}_S}_{u} + \underbrace{\alpha_{S^c}\mathbf{X}^T[\mathbf{I} - \alpha_S\mathbf{X}^T(\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T)^{-1}\mathbf{X}\alpha_S^T]\left(\frac{w}{\lambda_n n}\right)}_{v}.$$

By the triangle inequality, we have $\|\hat{z}_{S^c}\|_\infty \le \|u\|_\infty + \|v\|_\infty$. With the Mutual incoherence condition (A2), we have $\|u\|_\infty \le p$. By the choice of λ_n (7), we have $\|v\|_\infty \le \frac{1}{2}(1-p)$. Putting together this pieces, we have the strict dual feasibility condition holds, $\|\hat{z}_{S^c}\|_\infty < 1$.

It remains to establish a bound on the l_∞-norm of the error $\hat{\theta}_S - \theta_S^\star$. From equation (19), and triangle inequality, it holds that

$$\|\hat{\theta}_S - \theta_S^\star\|_\infty = \left\| \left(\frac{\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T}{n}\right)^{-1}\alpha_S\mathbf{X}^T\frac{w}{n}\right\|_\infty + \lambda_n \left\|\left(\frac{\alpha_S\mathbf{X}^T\mathbf{X}\alpha_S^T}{n}\right)^{-1}\right\|_\infty, \tag{19}$$

which completes the proof. □

Lemma 5 If the lower eigenvalue condition (A1) holds, then success of the PDW construction implies that the vector $(\hat{\theta}_S, 0) \in \mathbb{R}^m$ is the unique optimal solution of the RLKR program.

Proof When the PDW construction succeeds, then $(\hat{\theta}_S, 0)$ is an optimal solution with associated subgradient vector $\hat{z}\mathbb{R}^m$ satisfying $\|\hat{z}_{S^c}\|_\infty < 1$, and $< \hat{z}, \hat{\theta} >= \|\hat{\theta}\|_1$. Now let $\tilde{\theta}$ be any other optimal solution. If we introduce the shorthand notation $F(\theta) = \frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\alpha^T\theta\|_2^2$, then it holds that$F(\hat{\theta}) + \lambda_n < \hat{z}, \hat{\theta} >= F(\tilde{\theta}) + \lambda_n\|\tilde{\theta}\|_1$, and hence

$$F(\hat{\theta}) - \lambda_n < \hat{z}, \tilde{\theta} - \hat{\theta} >= F(\tilde{\theta}) + \lambda_n(\|\tilde{\theta}\|_1 - < \hat{z}, \tilde{\theta} >).$$

By the zero-subgradient conditions (15), we have $\lambda_n\hat{z} = -\nabla F(\hat{\theta})$, which implies that

$$F(\hat{\theta}) + < \nabla F(\hat{\theta}), \tilde{\theta} - \hat{\theta} > -F(\tilde{\theta}) = \lambda_n(\|\tilde{\theta}\|_1 - < \hat{z}, \tilde{\theta} >).$$

By the convex of $F(\theta)$, the left side is negative, which implies $\|\tilde{\theta}\|_1 \leq < \hat{z}, \tilde{\theta} >$. Combined with $< \hat{z}, \tilde{\theta} > \leq \|\hat{z}\|_\infty\|\tilde{\theta}\|_1$, we must have $\|\tilde{\theta}\|_1 =< \hat{z}, \tilde{\theta} >$. And for $\hat{z}_{S^c} < 1$, it follows that $\tilde{\theta}_{S^c} = 0$.

Thus, all optimal solutions are supported only on S , and hence can be obtained by solving the oracle subproblem (16). Given the lower eigenvalue condition (A1), this subproblem is strictly convex, and so has a unique minimizer. □

Proof **(Theorem 4)** Throughout the proof, we denote $\hat{\Delta} = \hat{\theta} - \theta^\star$.

(a) As $\hat{\theta}$ is the minimizer of the Regular Linear Kernel Regression program, we have

$$\frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\alpha^T\hat{\theta}\|_2^2 + \lambda_n\|\hat{\theta}\|_1 \leq \frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\alpha^T\theta^\star\|_2^2 + \lambda_n\|\theta^\star\|_1.$$

By some calculation, it follows that

$$0 \leq \frac{1}{2n}\|\mathbf{X}\hat{\Delta}\|_2^2 \leq \frac{w^T\mathbf{X}\alpha^T\hat{\Delta}}{n} + \lambda_n\{\|\theta^\star\|_1 - \|\hat{\theta}\|_1\}. \tag{20}$$

By Hölder inequality and choice of λ_n, we have

$$|\frac{w^T\mathbf{X}\alpha^T}{n}| \leq \left\|\frac{\alpha\mathbf{X}^Tw}{n}\right\|_\infty \|\hat{\Delta}\|_1 \leq \frac{\lambda_n}{2}\{\|\theta^\star\|_1 + \|\hat{\theta}\|_1\}.$$

Putting together the pieces yields

$$0 \leq \frac{\lambda_n}{2}\{\|\theta^\star\|_1 + \|\hat{\theta}\|_1\} + \lambda_n\{\|\theta^\star\|_1 - \|\hat{\theta}\|_1\},$$

which implies that $\|\hat{\theta}\|_1 \leq 3\|\theta^\star\|_1$. Consequently, by triangle inequality, we have $\|\hat{\Delta}\|_1 \leq \|\theta^\star\|_1 + \|\hat{\theta}\|_1 \leq 4\|\theta^\star\|_1$.

Returning to the early inequality (20), we have

$$\frac{1}{2n}\|\mathbf{X}\hat{\Delta}\|_2^2 \leq \frac{\lambda_n}{2}\|\hat{\Delta}\|_1 + \lambda_n\{\|\theta^\star\|_1 - \|\theta^\star - \hat{\Delta}\|_1\} \leq \frac{3\lambda_n}{2}\|\hat{\Delta}\|_1.$$

Combined with $\|\hat{\Delta}\|_1 \leq 4\|\theta^\star\|_1$, the proof is completed.

(b) With the same argument with the proof of (a), we have

$$\begin{aligned} 0 \leq \frac{1}{2n}\|\mathbf{X}\hat{\Delta}\|_2^2 &\leq \frac{\lambda_n}{2}\|\hat{\Delta}\|_1 + \lambda_n\{\|\theta^\star\|_1 - \|\hat{\theta}\|_1\} \\ &\overset{i}{\leq} \frac{\lambda_n}{2}\|\hat{\Delta}\|_1 + \lambda_n\{\|\hat{\Delta}_S\|_1 - \|\hat{\Delta}_{S^c}\|_1\} \\ &\leq \frac{\lambda_n}{2}\{3\|\hat{\Delta}_S\|_1 - \|\hat{\Delta}_{S^c}\|_1\}. \end{aligned} \tag{21}$$

The inequality (i) comes from the S-sparse of $\theta^\star$. Consequently, $\hat{\Delta} \in \mathbb{C}_3(S)$, by the the $(\kappa; 3)$-RE condition of $\mathbf{X}\alpha^T$, it holds that

$$\kappa\|\hat{\Delta}\|_2^2 \leq 3\lambda_n\|\hat{\Delta}_S\|_1 \leq 3\lambda_n\sqrt{s}\|\hat{\Delta}_S\|_2.$$

From the inequality (21), we have

$$\frac{1}{n}\|\mathbf{X}\hat{\Delta}\|_2^2 \leq 3\lambda_n\|\hat{\Delta}_S\|_1 \leq 3\lambda_n\sqrt{s}\|\hat{\Delta}_S\|_2.$$

Putting together this pieces, we have

$$\frac{1}{n}\|\mathbf{X}\hat{\Delta}\|_2^2 \leq \frac{9}{\kappa}s\lambda_n^2.$$

□

Proof **(Lemma 2)** Consider the moment generation function of zero-mean sub-Gaussian variable X,

$$\mathbb{E}\left[e^{\lambda X}\right] \leq e^{\frac{\sigma^2\lambda^2}{2}}.$$

By Hoeffding bound, we have

$$P\left[X \geq t\right] \leq e^{-\frac{t^2}{2\sigma^2}}.$$

From the definition of Z, it holds that

$$\begin{aligned} P[Z \geq t] &\leq \sum_{i=1}^{n}\{P[X_i \geq t] + P[-X_i \geq t]\} \\ &\leq 2ne^{-\frac{t^2}{2\sigma^2}}. \end{aligned}$$

□

References

[1] Fauci, A.S., Lane, H.C., Redfield, R.R.: Covid-19 - navigating the uncharted. New England Journal of Medicine **382**(13), 1268–1269 (2020) https://arxiv.org/abs/https://doi.org/10.1056/NEJMe2002387. https://doi.org/10.1056/NEJMe2002387. PMID: 32109011

[2] Yuki, K., Fujiogi, M., Koutsogiannaki, S.: Covid-19 pathophysiology: A review. Clinical immunology **215**, 108427 (2020)

[3] Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A.C., Hohmann, E., Chu, H.Y., Luetkemeyer, A., Kline, S., *et al.*: Remdesivir for the treatment of covid-19 - preliminary report. New England Journal of Medicine **383**(10), 992–994 (2020) https://arxiv.org/abs/https://doi.org/10.1056/NEJMc2022236. https://doi.org/10.1056/NEJMc2022236

[4] Felsenstein, S., Herbert, J.A., McNamara, P.S., Hedrich, C.M.: Covid-19: Immunology and treatment options. Clinical immunology **215**, 108448 (2020)

[5] Lotfi, M., Hamblin, M.R., Rezaei, N.: Covid-19: Transmission, prevention, and potential therapeutic opportunities. Clinica chimica acta **508**, 254–266 (2020)

Table 4 The risk assessment of all 189 countries or areas.

index	Area	*risk*	index	Area	*risk*
1	United States	17.62	47	Egypt	-1.29
2	India	13.6	48	Czechia	-1.29
3	Brazil	7.31	49	Georgia	-1.31
4	Russia	2.59	50	Portugal	-1.31
5	United Kingdom	2.49	51	Israel	-1.31
6	France	1.92	52	Austria	-1.32
7	Argentina	1.39	53	Saudi Arabia	-1.33
8	Colombia	1.15	54	China	-1.33
9	Mexico	0.98	55	Croatia	-1.34
10	Spain	0.98	56	Kenya	-1.34
11	Peru	0.87	57	Bosnia and Herzegovina	-1.35
12	Italy	0.69	58	Cameroon	-1.36
13	Germany	0.66	59	Serbia	-1.36
14	Iran	0.44	60	Afghanistan	-1.37
15	Indonesia	0.37	61	Kazakhstan	-1.37
16	Turkey	0.3	62	Nigeria	-1.37
17	South Africa	-0.18	63	Oman	-1.38
18	Poland	-0.18	64	Chile	-1.38
19	Ukraine	-0.25	65	Uganda	-1.38
20	Philippines	-0.54	66	Algeria	-1.39
21	Belgium	-0.55	67	Jordan	-1.39
22	Romania	-0.66	68	Zimbabwe	-1.39
23	Canada	-0.67	69	Zambia	-1.39
24	Netherlands	-0.9	70	Ireland	-1.4
25	Malaysia	-0.91	71	Azerbaijan	-1.4
26	Japan	-0.96	72	Namibia	-1.4
27	Vietnam	-0.99	73	Tanzania	-1.41
28	Bangladesh	-1.06	74	Democratic Republic of Congo	-1.41
29	Tunisia	-1.09	75	Angola	-1.41
30	Pakistan	-1.12	76	Bahrain	-1.41
31	Hungary	-1.14	77	Libya	-1.42
32	Nepal	-1.15	78	Sudan	-1.42
33	Sweden	-1.15	79	United Arab Emirates	-1.42
34	Ecuador	-1.16	80	Mozambique	-1.42
35	Bulgaria	-1.17	81	Lebanon	-1.42
36	Thailand	-1.17	82	North Macedonia	-1.43
37	Bolivia	-1.19	83	Moldova	-1.43
38	Greece	-1.19	84	Slovakia	-1.43
39	Iraq	-1.2	85	Malawi	-1.43
40	Pakistan	-1.21	86	Rwanda	-1.44
41	Paraguay	-1.21	87	Somalia	-1.44
42	Sri Lanka	-1.23	88	Kuwait	-1.44
43	Morocco	-1.23	89	Lithuania	-1.44
44	Switzerland	-1.24	90	Armenia	-1.44
45	Myanmar	-1.26	91	Syria	-1.44
46	Ethiopia	-1.28	92	Botswana	-1.45

[6] Ndaïrou, F., Area, I., Nieto, J.J., Torres, D.F.: Mathematical modeling of covid-19 transmission dynamics with a case study of wuhan. Chaos, Solitons & Fractals **135**, 109846 (2020)

[7] Gostic, K., Gomez, A.C., Mummah, R.O., Kucharski, A.J., Lloyd-Smith, J.O.: Estimated effectiveness of symptom and risk screening to prevent

Table 5 The risk assessment of all 189 countries or areas.

index	Area	*risk*	index	Area	*risk*
93	Belarus	-1.45	142	Liberia	-1.52
94	Yemen	-1.45	143	Senegal	-1.52
95	Burundi	-1.45	144	Vatican	-1.52
96	Comoros	-1.45	145	Honduras	-1.52
97	Maldives	-1.46	146	Sierra Leone	-1.52
98	South Korea	-1.46	147	Gibraltar	-1.53
99	Uruguay	-1.46	148	Mauritania	-1.53
100	Singapore	-1.46	149	Gambia	-1.53
101	Monaco	-1.46	150	Luxembourg	-1.53
102	Qatar	-1.46	151	Guinea-Bissau	-1.54
103	Denmark	-1.46	152	San Marino	-1.54
104	Slovenia	-1.46	153	Brunei	-1.54
105	South Sudan	-1.46	154	Estonia	-1.54
106	Eritrea	-1.46	155	Venezuela	-1.54
107	Seychelles	-1.46	156	Liechtenstein	-1.54
108	Uzbekistan	-1.46	157	Andorra	-1.55
109	Australia	-1.46	158	Martinique	-1.55
110	Mauritius	-1.46	159	Cuba	-1.55
111	Djibouti	-1.47	160	Timor	-1.56
112	Cambodia	-1.47	161	Cape Verde	-1.56
113	Chad	-1.47	162	Dominican Republic	-1.58
114	Lesotho	-1.47	163	Panama	-1.58
115	Norway	-1.47	164	Trinidad and Tobago	-1.59
116	Congo	-1.47	165	Papua New Guinea	-1.59
117	Niger	-1.47	166	Iceland	-1.6
118	Equatorial Guinea	-1.48	167	Suriname	-1.61
119	Central African Republic	-1.48	168	Guyana	-1.63
120	Gabon	-1.48	169	New Zealand	-1.64
121	Ghana	-1.49	170	New Caledonia	-1.64
122	Sao Tome and Principe	-1.49	171	Jamaica	-1.65
123	Cyprus	-1.49	172	Barbados	-1.65
124	Burkina Faso	-1.49	173	Saint Lucia	-1.65
125	Cote d'Ivoire	-1.49	174	Grenada	-1.65
126	Albania	-1.5	175	Antigua and Barbuda	-1.65
127	Benin	-1.5	176	Dominica	-1.66
128	Togo	-1.5	177	Saint Kitts and Nevis	-1.66
129	Mali	-1.5	178	Anguilla	-1.66
130	Tajikistan	-1.5	179	El Salvador	-1.66
131	Latvia	-1.5	180	Haiti	-1.66
132	Guatemala	-1.5	181	Montserrat	-1.66
133	Finland	-1.5	182	Bahamas	-1.67
134	Montenegro	-1.5	183	Aruba	-1.67
135	Costa Rica	-1.51	184	Turks and Caicos Islands	-1.67
136	Kyrgyzstan	-1.51	185	Fiji	-1.67
137	Mongolia	-1.51	186	Belize	-1.69
138	Malta	-1.51	187	Nicaragua	-1.69
139	Guinea	-1.51	188	Samoa	-1.7
140	Laos	-1.52	189	French Polynesia	-1.72
141	Bhutan	-1.52			

the spread of covid-19. Elife **9**, 55570 (2020)

[8] Bleier, B.S., Ramanathan Jr, M., Lane, A.P.: Covid-19 vaccines may not prevent nasal sars-cov-2 infection and asymptomatic transmission. Otolaryngology–Head and Neck Surgery **164**(2), 305–307 (2021)

[9] Osborne, M.R., Presnell, B., Turlach, B.A.: On the lasso and its dual. Journal of Computational and Graphical statistics **9**(2), 319–337 (2000)

[10] Zhao, P., Yu, B.: On model selection consistency of lasso. The Journal of Machine Learning Research **7**, 2541–2563 (2006)

[11] Hastie, T., Montanari, A., Rosset, S., Tibshirani, R.J.: Surprises in High-Dimensional Ridgeless Least Squares Interpolation (2020)

[12] Bartlett, P.L., Long, P.M., Lugosi, G., Tsigler, A.: Benign overfitting in linear regression. Proceedings of the National Academy of Sciences **117**(48), 30063–30070 (2020) https://arxiv.org/abs/https://www.pnas.org/content/117/48/30063.full.pdf. https://doi.org/10.1073/pnas.1907378117

[13] Tuo, R., Wang, Y., Jeff Wu, C.F.: On the improved rates of convergence for matérn-type kernel ridge regression with application to calibration of computer models. SIAM/ASA Journal on Uncertainty Quantification **8**(4), 1522–1547 (2020) https://arxiv.org/abs/https://doi.org/10.1137/19M1304222. https://doi.org/10.1137/19M1304222

[14] Liu, F., Liao, Z., Suykens, J.: Kernel regression in high dimensions: Refined analysis beyond double descent. In: Banerjee, A., Fukumizu, K. (eds.) Proceedings of The 24th International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research, vol. 130, pp. 649–657. PMLR, ??? (2021). https://proceedings.mlr.press/v130/liu21b.html

[15] Crawford, L., Wood, K.C., Zhou, X., Mukherjee, S.: Bayesian approximate kernel regression with variable selection. Journal of the American Statistical Association **113**(524), 1710–1721 (2018) https://arxiv.org/abs/https://doi.org/10.1080/01621459.2017.1361830. https://doi.org/10.1080/01621459.2017.1361830. PMID: 30799887

[16] Liang, T., Rakhlin, A.: Just interpolate: Kernel "ridgeless" regression can generalize. The Annals of Statistics **48**(3), 1329–1347 (2020)

[17] Braga, I., Monard, M.C.: Improving the kernel regularized least squares method for small-sample regression. Neurocomputing **163**, 106–114 (2015)

Figures

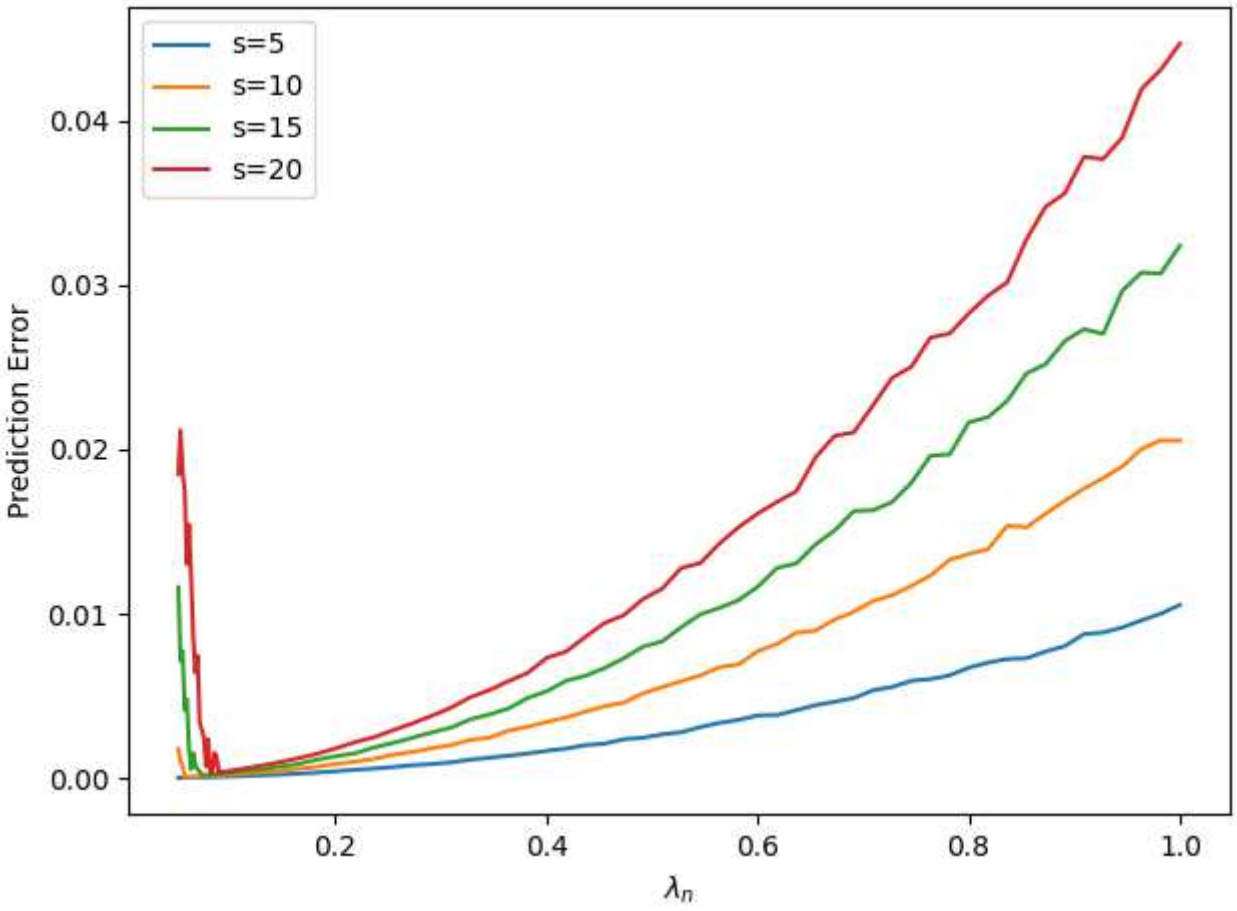

Figure 1

The Prediction error of Regular Linear Kernel Regression.