ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

Measuring Dependence With HSIC

Chengyu Du[1], Yingtong Peng[2], Zhengyuan Qu[3], Chixueyuan Wang[4]

[1]School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, China.

[2]College of Liberal Arts, University of Minnesota - Twin Cities, Minneapolis, MN 55455, USA.

[3]Faculty of Science and Engineering, University of Nottingham Ningbo China, Ningbo, 315100, China.

[4]Faculty of Arts and Science, University of Toronto, Toronto, M5B 0A5, Canada.

Abstract

This paper proposes a method of nonlinear independence test between random variables. The method is a kernel method called Hilbert–Schmidt independence criterion (HSIC) for measuring dependence, which is based on the Hilbert– Schmidt norm of the cross-covariance operator of samples. Unlike previous methods like mutual information, it allows us to simplify the problem and to deal with several sensitive variables in the mean time. And it uses a kernel function instead of linear formulas, which makes it possible to solve nonlinear problems. This work studies the detailed concepts and criterion associated with HSIC and use the median heuristic To select the kernel width. The experiments are conducted on the wine dataset in UCI and the results showed good performance.

Keywords

Independence test, HSIC, median heuristic.

1. INTRODUCTION

1.1. Motivation

With the advances in technology, machine learning algorithms have developed rapidly. At present, the widespread applications of machine learning algorithms have also had a tremendous impact on people's daily life. For example, they can influence and predict some weather-related policies, as well as economic and educational decisions. Large amounts of big data and mechanical learning algorithms are now being exploiting and utilizing for prediction and inference. For instance, algorithms are now used to determine wages [1], help assess before trial [2], and assess the risk of violence [3]. It seems that because of society's strong reliance on machine learning algorithms, companies and government agencies are beginning to concern about the fairness and impartiality of machine learning. Indeed, standard machine learning models are not fair in any way, because they retain biases in the data, and human-provided labels is highly likely to be affected by cognitive biases. So, more meticulous modeling are needed to get closer to more equitable machine learning algorithms.

This paper uses a novel fair regression based on the fair classification framework mentioned in [4]. This method relies on the Hilbert Schmidt independence criterion as the fairness term. This paper is based on two literatures [4, 5]. Among them, the independence between the predictor and the sensitive variable is achieved by using the kernel dependence measure (HSIC), as the regularizer of the objective function [5]. This work is to study the application of HSIC in fairness and show how it can be reflected through a comparison test with and without median heuristic. Firstly, there is an introduction of the expressions and derivation process of HSIC and

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

median heuristic, and the framework of fairness regression [4]. After that, there is a database to verify the function of HSIC, thus proving that HSIC can realize fairness in statistical machine learning.

1.2. Benefits of HSIC and Median Heuristic

The classical method of measuring dependencies between variables is usually designed to capture only a specific form of dependency and is applicable only to scalar random variables. For instance, that mutual information can be used to measure the independence between two variables, and when it is equal to zero it means that the two variables are independent of each other. But the adoption of mutual information means that it is essential to estimate some probability density. Obviously, many problems cannot estimate the probability or probability density very well, and the two variables are likely to have very different meanings. A new and more general framework was introduced in [6], which is called Hilbert Schmidt Independence Criterion (HSIC) and based on calculating the distance between the probabilistic measure representations using kernels. It can capture all possible types of dependence between random vectors or random variables valued in the non-Euclidean or structured domains. HSIC can be skillfully transformed into a sampling form, avoiding the need to estimate the probability density. Its main purpose is to measure a distribution difference between two variables, which is similar to covariance (variance) and is also constructed based on covariance. There is a theorem for HSIC: if and only if the variables are independent, HSIC will be equal to zero.

According to the content mentioned in [6], HSIC has many advantages over previous kernel independence measures: the empirical estimate is simpler; HSIC-based independence test will not have the problem of slow learning speed; the finite sample bias of estimates can be small enough to be negligible; it is superior to previous independent tests and has a strong competitiveness.

As is known to all, the performance of kernel methods largely depends on the choice of kernel. However, according to article [7], there is no general, theoretical basis for kernel selection in this case, because it is sometimes not possible to select kernel parameters using a simple cross-validation method, as opposed to a supervised kernel method, therefore, an ideal objective function cannot be calculated due to a complex asymptotic null distribution. Therefore, most users of kernel hypothesis testing resort to parameterized kernel families, such as squared exponents, and set the length scale parameters based on the "median heuristic". As [7] mentioned, there are two versions of expression of median heuristic. The first one uses Gaussian RBF and the expression is $k(x, x') = exp\left(-\frac{||x-x'||^2}{l^2}\right)$, and the second version uses the parameterization $k(x, x') = exp\left(-\frac{||x-x'||^2}{2l^2}\right)$. In both versions, given a set of observations $x_1, \dots. x_n$, this work calculates l = median $\left(||x_i - x_j||_2\right)$. The median heuristic is utilized as the default value for supervised learning tasks, or when cross-validation is too expensive.

1.3. Fairness in Machine Learning

Obtaining fair machine learning algorithms is an urgent problem. Fairness is an elusive and unpredictable concept, so such qualitative indicators are contained in machines that only learn from data. Therefore, in order to achieve the fairness of machine learning, there is a goal: reduce the weight of sensitive features or directly remove them.

However, removing sensitive characteristics such as gender and race to make predictions does not solve the problem, since the accuracy of the model may be greatly affected by the lack of information features, and related variables may still enter the model [4]. Based on this

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

premise, we derive the HSIC method. When modifying the trade-off between prediction and fairness, it allows us to study different solutions. The method can naturally handle multiple dimensional input variables for both regular and sensitive variables.

In this paper, we have not described too much about the use of fairness and algorithms. The focus of our research is on HSIC in fairness. The remainder of this article is organized as follows. Section 2 introduces the basic expressions of HSIC, and gives a framework for median heuristic. In Section 3, we select a data set and conducted a comparative experiment and gives experimental evidence and results. Finally, this paper is summarized in Section 4.

2. METHODOLOGY

This section introduces a criterion for judging the independence between two random variables based on the RKHS(Reproducing Kernel Hilbert Spaces).2.1 will introduce operators used by HSIC(The Hilbert-Schmidt Independence Criterion), then 2.2 will give the definition of HSIC, and finally, 2.3 will explain a method of parameter selection: Median heuristic

2.1. The Cross-Covariance Operator

Firstly, the work presents the cross-covariance operator in linear cases. To make it clear, here defining some notations. Two spaces $\mathcal{X} \subseteq R$ and $\mathcal{Y} \subseteq R$, which consisting of sample observation pairs (x, y) of bivariate distribution P_{xy} that we collect are defined. Then the cross-covariance operator can be written as [8]:

$$C_{xy} := \mathbf{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^\top] - \mathbf{E}_{\mathbf{x}}[\mathbf{x}]\mathbf{E}_{\mathbf{y}}[\mathbf{y}^\top] \tag{1}$$

Where $\mathbf{y}^\top$ is the transpose of $\mathbf{y}$, where $\mathbf{E}_{x,y}$ is the expection of P_{xy}, where $\mathbf{E}_{\mathbf{x}}$ and $\mathbf{E}_{\mathbf{y}}$ are the expections of P_x and P_y respectively.

Secondly, this work presents it in non-linear cases. To do this, kernel functions need to be used. We define a reproducing kernel Hilbert space $\mathcal{F}$ of functions from $\mathcal{X} \subseteq R$, there correspond an element $\phi(x) \in \mathcal{F}$ such that $\langle \phi(x), \phi(x') \rangle_{\mathcal{F}} = k_x(x, x')$, where $k(\cdot,\cdot)$ is a positive definite kernel function [9]. The other suitable positive definite kernel function $k_y(y, y') = \langle \psi(y), \psi(y') \rangle$ of another reproducing kernel Hilbert space $\mathcal{G}$ also be defined. Then the cross-covariance operator can be presented like (1) which is a linear operator $C_{xy}: \mathcal{G} \to \mathcal{F}$:

$$C_{xy} := \mathbf{E}_{x,y}[(\phi(x) - \mu_x) \otimes (\psi(y) - \mu_y)]$$

Where $\otimes$ is the notation of tensor product, $\mu_x = E_x[\phi(x)]$ and $\mu_y = E_y[\psi(y)]$.

2.2. The Hilbert-Schmidt Independence Criterion (HSIC)

The squared HS-norm cross-covariance operator C_{xy} is defined as The Hilbert-Schmidt Independence Criterion (HSIC) [8] such that:

$$\begin{aligned} HSIC(\mathbb{P}_{xy}, \mathcal{F}, \mathcal{G}) &= \| C_{xy} \|_{\mathrm{HS}}^2 \\ &= \mathbb{E}_{xyx'y'}[k_x(x, x')k_y(y, y')] \\ &+ \mathbb{E}_{xx'}[k_x(x, x')]\mathbb{E}_{yy'}[k_y(y, y')] \\ &- 2\mathbb{E}_{xy}[\mathbb{E}_{x'}[k_x(x, x')]\mathbb{E}_{y'}[k_y(y, y')]] \end{aligned}$$

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

Where $E_{xyx'y'}$ represents the independent expectation of (x, y) and (x', y') derived from P_{xy}, where k_x and k_y are kernel functions, where $\mathcal{F}$ and $\mathcal{G}$ are the Hilbert spaces, and where $\mathbb{E}_{xy}$ is the expectation over X and Y.

By defining a data series $Z := \{(x_1, y_1), \dots, (x_m, y_m)\}$ made up of m independent observations pairs drawn from P_{xy} with $\mathcal{X} \subseteq R$ and $\mathcal{Y} \subseteq R$ The empirical estimator of HSIC can be written as:

$$HSIC(\mathcal{F}, \mathcal{G}, \mathcal{Z}) = (m-1)^{-2}\mathrm{tr}(\mathbf{K}_X \mathbf{H} \mathbf{K}_Y \mathbf{H})$$

Where m is the number of samples, $\mathbf{K}_X \in R^{m \times m}$ and $\mathbf{K}_Y \in R^{m \times m}$ with entries are $\mathbf{K}_{Xij} = k(x_i, x_j)$ and $\mathbf{K}_{Yij} = k(y_i, y_j)$, and $\mathbf{H} \in R^{m \times m}$ is the centering matrix $H = I_m - \frac{1}{m}\mathbf{1}_m\mathbf{1}_m^T$.

2.3. Median Heuristic

It is well known that the key to the efficiency and success of kernel methods lies in kernel function selection. In practice, this choice tends to be reduced to bandwidth ν calibration and may even be more important than kernel selection. For instance, assume that we choose Gaussian kernel. When $\nu \to 0$, the Gram matrix K is the identity matrix, and when $\nu \to \infty$, everything in K is 1 [9]. In these two extreme cases, the information contained in the data cannot be represented. This is a universal situation, hence choosing a value in the middle area $(\| x_i - x_j \|)_{1 \le i, j \le n}$ is considered the correct middle choice. The median is usually set as the empirical quantile. This method is called median heuristic. This report chooses the bandwidth $\nu = \sqrt{H_n/2}$, where

$$H_n = \mathrm{Med}\{\| X_{n,i} - X_{n,j} \|^2 | \ 1 \le i < j \le n\}$$

Where Med is the empirical median

More details can be found in [9], if readers are interested.

3. EXPERIMENTAL EVIDENCE

3.1. Dataset

A dataset is chosen from the UCI repository: the Wine dataset. These data are the chemical composition analysis of three different wines grown in Italy, including the content of thirteen chemical components.

3.2. Experimental Set

The wine data is divided into three groups according to the type. First, when measuring dependence, the work needs to specify the kernels for X and Y. The Gaussian kernels is used and default value of the width parameter is 1.0. Theoretically, the selection of the HSIC kernel function does not affect the test results.

This work measured the dependence of 13 attributes for each group and use the following symbols to represent these 13 variables.

A=Alcohol;

B=Malic acid;

C=Ash;

D=Alkalinity of ash;

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

E=Magnesium;
F=Total phenols;
G=Flavonoids;
H=Nonflavonoid phenols;
I=Proanthocyanins;
J=Color intensity;
K=Hue;
L=OD280/OD315 of diluted wines;
M=Proline;

3.3. Result

The significance level is set to 0.01. For the convenience of statistical results, it is set to 0 when the p-value is less than 0.01, thinking that it rejects the null hypothesis. Table 1, 2 and 3 show the results.

Table 1. TypeA

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0.00E+00	8.34E-01	1.94E-01	2.70E-01	6.78E-01	0.00E+00	0.00E+00	9.28E-01	0.00E+00	0.00E+00	7.29E-01	4.91E-01	2.03E-01
B	8.78E-01	0.00E+00	5.49E-01	1.43E-01	1.71E-01	9.07E-01	2.24E-01	3.36E-01	5.64E-01	4.10E-02	0.00E+00	2.39E-01	6.25E-01
C	2.21E-01	3.08E-01	0.00E+00	1.19E-01	8.41E-01	7.99E-01	4.99E-01	0.00E+00	2.77E-01	7.20E-01	7.29E-02	5.50E-01	3.90E-02
D	1.38E-01	1.75E-01	1.52E-01	0.00E+00	8.09E-01	1.11E-01	3.14E-01	3.63E-01	7.79E-02	6.97E-01	4.59E-01	3.71E-01	2.58E-01
E	7.85E-01	6.88E-01	7.57E-01	4.33E-01	0.00E+00	7.79E-02	1.60E-02	5.22E-01	5.19E-01	7.39E-02	8.57E-01	1.53E-01	3.36E-01
F	0.00E+00	7.33E-01	8.35E-01	1.40E-01	7.29E-02	0.00E+00	0.00E+00	9.26E-01	0.00E+00	0.00E+00	1.04E-01	6.74E-01	6.05E-01
G	0.00E+00	1.99E-01	5.38E-01	1.04E-01	1.40E-02	0.00E+00	0.00E+00	5.00E-01	0.00E+00	0.00E+00	9.02E-01	5.17E-01	6.09E-02
H	8.73E-01	3.28E-01	0.00E+00	2.04E-01	4.75E-01	8.70E-01	3.91E-01	0.00E+00	2.89E-01	3.03E-01	0.00E+00	1.60E-02	2.46E-01
I	0.00E+00	6.00E-01	2.87E-01	1.09E-01	8.19E-02	0.00E+00	0.00E+00	2.99E-01	0.00E+00	0.00E+00	4.63E-01	6.46E-01	4.31E-01
J	0.00E+00	1.40E-02	8.57E-01	6.97E-01	4.35E-01	0.00E+00	0.00E+00	2.73E-01	0.00E+00	0.00E+00	7.06E-01	1.40E-01	7.10E-01
K	7.24E-01	0.00E+00	7.49E-02	5.78E-01	9.29E-01	1.11E-01	9.20E-01	0.00E+00	4.23E-01	6.64E-01	0.00E+00	1.20E-02	2.81E-01
L	5.59E-01	2.04E-01	5.41E-01	2.85E-01	1.68E-01	8.27E-01	1.90E-02	1.90E-02	6.14E-01	6.29E-02	2.00E-02	0.00E+00	1.97E-01
M	5.67E-01	3.17E-01	2.46E-01	4.31E-01	5.11E-01	7.93E-01	8.66E-01	3.97E-01	2.36E-01	9.17E-01	8.67E-01	2.18E-01	0.00E+00

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

Table 2. TypeB

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0.00E+00	2.36E-01	1.80E-02	7.89E-02	9.37E-01	7.91E-01	8.52E-01	9.05E-01	3.72E-01	2.30E-02	5.94E-01	3.37E-01	8.44E-01
B	3.57E-01	0.00E+00	2.21E-01	7.92E-01	9.57E-01	6.84E-01	8.18E-01	1.76E-01	1.33E-01	2.77E-01	0.00E+00	5.75E-01	7.09E-02
C	3.40E-02	2.24E-01	0.00E+00	0.00E+00	2.68E-01	5.12E-01	2.72E-01	1.70E-02	4.73E-01	3.52E-01	9.23E-01	4.03E-01	5.89E-02
D	4.28E-01	2.88E-01	0.00E+00	0.00E+00	5.99E-03	7.87E-01	5.04E-01	5.89E-02	2.82E-01	4.21E-01	1.20E-02	3.50E-02	2.35E-01
E	4.47E-01	7.47E-01	5.53E-01	3.69E-01	0.00E+00	8.99E-02	6.85E-01	3.34E-01	4.60E-01	1.18E-01	3.87E-01	4.45E-01	7.69E-02
F	4.95E-01	7.32E-01	4.94E-01	3.37E-01	7.55E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.24E-01	8.32E-01	0.00E+00	1.68E-01
G	9.47E-01	6.82E-01	2.08E-01	4.95E-01	2.79E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.60E-02	8.54E-01	0.00E+00	4.70E-01
H	7.70E-01	1.53E-01	1.40E-02	6.29E-02	1.30E-01	0.00E+00	0.00E+00	0.00E+00	2.10E-02	8.06E-01	7.52E-01	0.00E+00	2.80E-01
I	6.57E-01	2.23E-01	2.34E-01	3.11E-01	1.82E-01	0.00E+00	0.00E+00	3.00E-02	0.00E+00	9.06E-01	9.38E-01	0.00E+00	1.83E-01
J	1.90E-02	3.33E-01	2.89E-01	2.80E-01	1.36E-01	2.29E-01	3.01E-01	7.83E-01	8.91E-01	0.00E+00	8.85E-01	2.84E-01	7.57E-01
K	5.71E-01	0.00E+00	9.03E-01	5.59E-02	3.98E-01	8.11E-01	8.67E-01	6.75E-01	9.86E-01	8.33E-01	0.00E+00	6.95E-01	6.84E-01
L	3.59E-01	6.74E-01	4.85E-01	7.89E-02	8.16E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.70E-01	6.83E-01	0.00E+00	9.04E-01
M	8.75E-01	3.30E-02	5.13E-01	5.19E-01	8.69E-02	8.88E-01	6.74E-01	1.43E-01	7.30E-01	2.75E-01	9.79E-02	6.13E-01	0.00E+00

Table 3. TypeC

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0.00E+00	6.99E-01	2.44E-01	3.03E-01	3.96E-01	2.68E-01	8.29E-01	6.34E-01	2.60E-02	1.04E-01	8.80E-01	3.91E-01	5.38E-01
B	6.41E-01	0.00E+00	8.47E-01	3.49E-01	3.15E-01	2.93E-01	2.18E-01	5.92E-01	3.20E-02	5.79E-01	6.21E-01	5.56E-01	5.64E-01
C	1.36E-01	9.49E-01	0.00E+00	0.00E+00	6.99E-02	0.00E+00	5.49E-02	9.67E-01	2.04E-01	7.51E-01	2.42E-01	1.11E-01	4.51E-01
D	1.41E-01	3.27E-01	0.00E+00	0.00E+00	6.75E-01	7.39E-02	5.09E-02	9.82E-01	2.86E-01	6.62E-01	6.80E-01	9.71E-01	1.23E-01
E	8.56E-01	4.69E-01	2.88E-01	4.41E-01	0.00E+00	7.39E-01	8.39E-02	3.60E-02	4.99E-01	5.21E-01	7.31E-01	5.59E-01	8.78E-01
F	2.09E-01	3.32E-01	0.00E+00	9.19E-02	5.31E-01	0.00E+00	8.89E-02	3.20E-02	0.00E+00	1.14E-01	5.89E-02	1.44E-01	7.67E-01
G	8.16E-01	2.20E-01	6.29E-02	6.19E-02	8.59E-02	9.09E-02	0.00E+00	0.00E+00	2.20E-02	1.06E-01	2.90E-01	0.00E+00	7.88E-01
H	6.82E-01	4.79E-01	9.73E-01	9.63E-01	9.69E-02	3.30E-02	0.00E+00	0.00E+00	2.95E-01	8.02E-01	0.00E+00	3.00E-02	6.17E-01
I	3.20E-02	1.46E-01	2.58E-01	1.12E-01	9.15E-01	0.00E+00	2.10E-02	2.88E-01	0.00E+00	0.00E+00	0.00E+00	1.14E-01	1.19E-01
J	3.94E-01	2.44E-01	8.27E-01	9.72E-01	3.04E-01	3.88E-01	6.44E-01	8.82E-01	0.00E+00	0.00E+00	0.00E+00	3.22E-01	9.10E-01
K	8.86E-01	7.22E-01	2.09E-01	6.10E-01	3.06E-01	6.45E-01	4.40E-02	0.00E+00	0.00E+00	0.00E+00	0.00E+00	1.10E-02	5.76E-01
L	4.20E-01	5.17E-01	1.05E-01	9.74E-01	5.61E-01	2.91E-01	0.00E+00	3.40E-02	2.52E-01	2.81E-01	1.20E-02	0.00E+00	7.97E-01
M	5.05E-01	6.18E-01	6.80E-01	7.89E-02	4.15E-01	3.15E-01	8.22E-01	5.70E-01	5.49E-02	6.90E-01	4.65E-01	3.66E-01	0.00E+00

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

Table 4. TypeA with median heuristic

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0.00E+00	5.49E-01	0.00E+00	1.70E-01	2.99E-01	0.00E+00	3.40E-02	2.77E-01	0.00E+00	0.00E+00	6.16E-01	2.29E-01	1.59E-01
B	7.87E-01	0.00E+00	9.02E-01	4.84E-01	1.72E-01	8.93E-01	8.27E-01	3.23E-01	8.63E-01	5.74E-01	4.10E-01	5.51E-01	1.74E-01
C	0.00E+00	2.84E-01	0.00E+00	2.10E-02	2.80E-01	7.14E-01	5.38E-01	3.70E-02	9.09E-02	7.55E-01	8.64E-01	6.97E-01	1.90E-02
D	3.03E-01	4.28E-01	2.10E-02	0.00E+00	8.37E-01	2.03E-01	2.78E-01	4.96E-01	2.63E-01	8.28E-01	3.22E-01	7.81E-01	9.85E-01
E	1.12E-01	6.29E-01	3.48E-01	6.44E-01	0.00E+00	3.60E-02	2.65E-01	1.60E-02	7.29E-02	1.60E-02	2.95E-01	2.07E-01	8.79E-02
F	0.00E+00	9.31E-01	3.32E-01	1.80E-01	1.90E-02	0.00E+00	0.00E+00	2.10E-01	5.79E-02	0.00E+00	1.07E-01	5.68E-01	6.44E-01
G	1.70E-02	8.49E-01	4.86E-01	5.79E-01	1.26E-01	0.00E+00	0.00E+00	5.49E-02	0.00E+00	0.00E+00	3.09E-01	7.49E-02	2.77E-01
H	3.10E-01	4.59E-01	1.90E-02	6.30E-01	1.90E-02	1.19E-01	9.69E-02	0.00E+00	1.69E-01	5.77E-01	0.00E+00	3.90E-02	4.56E-01
I	0.00E+00	8.75E-01	1.30E-01	1.07E-01	4.64E-01	2.00E-02	0.00E+00	3.16E-01	0.00E+00	0.00E+00	6.18E-01	2.83E-01	4.21E-01
J	0.00E+00	3.17E-01	2.68E-01	6.94E-01	3.70E-02	0.00E+00	0.00E+00	1.77E-01	0.00E+00	0.00E+00	5.29E-01	8.79E-02	2.10E-02
K	6.41E-01	4.75E-01	8.67E-01	2.04E-01	4.17E-01	6.39E-02	3.39E-01	0.00E+00	6.84E-01	5.79E-01	0.00E+00	0.00E+00	1.68E-01
L	1.78E-01	2.74E-01	8.65E-01	5.60E-01	2.41E-01	8.00E-01	3.63E-01	1.80E-02	8.18E-01	1.86E-01	0.00E+00	0.00E+00	0.00E+00
M	5.21E-01	4.41E-01	1.61E-01	4.88E-01	1.60E-01	8.05E-01	5.09E-02	4.70E-01	4.42E-01	1.40E-02	2.88E-01	0.00E+00	0.00E+00

Table 5. TypeB with median heuristic

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0.00E+00	1.81E-01	5.99E-02	5.04E-01	8.63E-01	5.02E-01	7.30E-01	1.98E-01	4.62E-01	2.71E-01	3.89E-01	8.52E-01	5.66E-01
B	1.40E-01	0.00E+00	5.49E-01	8.89E-02	8.52E-01	2.80E-01	7.23E-01	8.07E-01	1.35E-01	3.95E-01	3.40E-02	9.78E-01	5.76E-01
C	2.10E-02	3.25E-01	0.00E+00	0.00E+00	3.06E-01	6.85E-01	7.11E-01	1.40E-02	6.59E-01	4.35E-01	4.93E-01	5.76E-01	7.72E-01
D	4.62E-01	2.00E-01	0.00E+00	0.00E+00	9.09E-02	2.18E-01	6.10E-01	6.69E-02	0.00E+00	1.64E-01	0.00E+00	3.54E-01	4.21E-01
E	7.86E-01	7.10E-01	3.93E-01	6.89E-02	0.00E+00	8.18E-01	4.65E-01	4.23E-01	1.47E-01	3.07E-01	4.69E-01	3.55E-01	8.89E-02
F	3.26E-01	4.51E-01	9.30E-01	3.44E-01	7.25E-01	0.00E+00	0.00E+00	6.09E-02	0.00E+00	1.46E-01	5.51E-01	0.00E+00	1.30E-02
G	9.18E-01	9.07E-01	5.57E-01	4.69E-01	3.01E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	6.21E-01	7.29E-02	0.00E+00	3.80E-02
H	5.55E-01	7.84E-01	1.20E-02	1.39E-01	5.07E-01	3.10E-02	0.00E+00	0.00E+00	5.67E-01	3.32E-01	1.97E-01	0.00E+00	5.42E-01
I	3.34E-01	3.20E-02	5.04E-01	0.00E+00	1.03E-01	0.00E+00	0.00E+00	3.08E-01	0.00E+00	5.97E-01	8.68E-01	3.90E-02	8.57E-01
J	6.39E-02	4.94E-01	3.82E-01	8.39E-02	8.02E-01	1.37E-01	7.68E-01	3.13E-01	3.07E-01	0.00E+00	1.22E-01	4.38E-01	3.06E-01
K	8.24E-01	1.10E-02	4.87E-01	0.00E+00	5.83E-01	6.30E-01	5.75E-01	1.63E-01	8.26E-01	1.21E-01	0.00E+00	9.86E-01	3.29E-01
L	9.22E-01	8.34E-01	8.11E-01	7.16E-01	4.64E-01	0.00E+00	0.00E+00	0.00E+00	1.78E-01	2.24E-01	9.54E-01	0.00E+00	5.13E-01
M	5.21E-01	4.59E-01	5.10E-01	5.66E-01	2.90E-02	8.88E-01	9.89E-02	6.19E-01	8.05E-01	4.45E-01	2.98E-01	6.95E-01	0.00E+00

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

Theoretically, the values of the symmetrical positions in the above table should be equal because they belong to the test between two identical variables. However, the results show that they are not symmetrical. In order to study whether the kernel width selection will affect the difference in the p-value of the symmetrical position, the median heuristic method is used to test. Table 4, 5 and 6 show the results.

Table 6. TypeC with median heuristic

	A	B	C	D	E	F	G	H	I	J	K	L	M
A	0.00E+00	5.83E-01	8.33E-01	5.05E-01	5.37E-01	2.33E-01	3.67E-01	5.49E-02	8.39E-02	1.76E-01	6.59E-01	2.28E-01	8.09E-01
B	4.40E-01	0.00E+00	7.32E-01	7.56E-01	1.80E-01	2.40E-01	2.78E-01	5.09E-01	4.57E-01	6.19E-01	8.58E-01	9.54E-01	7.32E-01
C	9.03E-01	8.22E-01	0.00E+00	0.00E+00	1.88E-01	1.61E-01	3.38E-01	1.09E-01	5.28E-01	9.12E-01	2.08E-01	4.64E-01	8.04E-01
D	2.44E-01	7.70E-01	0.00E+00	0.00E+00	2.89E-01	3.24E-01	2.66E-01	2.85E-01	3.74E-01	8.36E-01	7.42E-01	4.72E-01	5.72E-01
E	1.49E-01	1.54E-01	1.78E-01	2.22E-01	0.00E+00	3.24E-01	2.60E-02	4.60E-02	5.25E-01	3.62E-01	4.25E-01	1.31E-01	4.35E-01
F	3.09E-01	2.51E-01	2.09E-01	2.82E-01	6.62E-01	0.00E+00	1.40E-02	1.01E-01	2.78E-01	3.34E-01	2.06E-01	7.10E-01	1.22E-01
G	5.49E-02	2.31E-01	7.13E-01	7.89E-02	2.50E-02	4.70E-02	0.00E+00	0.00E+00	9.19E-02	1.40E-02	5.40E-01	0.00E+00	2.90E-02
H	7.79E-02	8.23E-01	2.96E-01	4.79E-01	9.69E-02	3.30E-02	0.00E+00	0.00E+00	8.36E-01	6.80E-01	7.99E-02	0.00E+00	3.22E-01
I	2.90E-02	1.87E-01	7.53E-01	4.04E-01	5.09E-01	2.05E-01	1.10E-02	9.15E-01	0.00E+00	0.00E+00	2.60E-02	0.00E+00	3.80E-02
J	3.74E-01	6.41E-01	8.34E-01	8.77E-01	6.05E-01	3.97E-01	3.90E-02	8.26E-01	0.00E+00	0.00E+00	0.00E+00	1.40E-02	5.83E-01
K	4.26E-01	7.62E-01	4.27E-01	5.41E-01	4.66E-01	2.48E-01	4.40E-02	1.04E-01	1.40E-02	0.00E+00	0.00E+00	1.38E-01	1.24E-01
L	2.27E-01	9.55E-01	4.37E-01	8.23E-01	1.69E-01	3.44E-01	0.00E+00	0.00E+00	0.00E+00	2.70E-02	4.80E-02	0.00E+00	3.78E-01
M	8.56E-01	6.96E-01	9.69E-01	3.52E-01	8.55E-01	1.53E-01	3.39E-01	2.58E-01	8.79E-02	5.11E-01	5.39E-02	4.17E-01	0.00E+00

3.4. Analysist

In fact, although the P values of the symmetric positions are not exactly the same, they are of the same order of magnitude. At the same time, the rejection of the null hypothesis is symmetric, which means that this does not significantly affect the results of the experiment. However, due to the randomness of the P value in an order of magnitude, if choosing a different significance level, such as 0.05, it may produce different results. When using the median heuristic method to conduct experiments for many times, it can be found that the P value is more stable, which reduces the influence of extreme values on data statistics to a certain extent. Therefore, it is feasible to apply median heuristic to the kernel method.

4. CONCLUSION

When measuring the independence between variables, the basic idea is to calculate mutual information, but in many problems it is difficult to estimate the probability or probability density well. One possible solution is to transform it into a dual problem, using a similar confrontational thinking to learn mutual information (the idea of infomax), but this method may be unstable and is affected by the sampling scheme. The best solution is to have an indicator similar to "correlation coefficient", so that we can explicitly calculate and optimize this indicator.

ISSN: 2472-3703 DOI: 10.6911/WSRJ.202010_6(10).0013

So this report proposes a new kernel method for measuring the independence between various variables, which is based on HSIC and has good theoretical and practical performance. This work tested this method on the wine dataset and showed good performance. It greatly simplifies the calculation of empirical estimates, and can display statistical information that helps data inference. In the future, our work aims to study conditional independence testing, and then extend it to other machine learning algorithms, such as fair learning and neural networks.

ACKNOWLEDGEMENTS

The authors sincerely thank Dr.Dino Sejdinovic for his guidance and help.

REFERENCES

[1] Dieterich, W., Mendoza, C., Brennan, T.: : COMPAS risk scales: Demonstrating accuracy equity and predictive parity. Working paper, Northpointe Inc. Res. Dep. (2016) https://njoselson.github.io/pdfs/ ProPublica_Commentary_Final_070616.pdf

[2] Brennan, Tim, W.D., Ehret, B. (2009) Evaluating the predictive validity of the compas risk and needs assessment system. Criminal Justice and Beh. 36(1): 21–40

[3] Cunningham, M.D., Sorensen, J.R. (2006) Actuarial models for assessing prison violence risk. Assessment 13(3): 253–265

[4] P´erez-Suay, A., Laparra, V., Mateo-Garc´ıa, G., Mu˜noz-Mar´ı, J., G´omezChova, L., Camps-Valls, G. (2017) Fair Kernel Learning. https:// arxiv.org/pdf/1710.05578.pdf

[5] Li, Z., P´erez-Suay, A., Camps-Valls, G., Sejdinovic, D. (2019) Kernel Dependence Regularizers and Gaussian Processes with Applications to Algorithmic Fairness .https: // arxiv.org /pdf/1911.04322.pdf

[6] Gretton, A., Bousquet, O., Smola, A., Sch¨olkopf, B. (2005) Measuring Statistical Dependence with Hilbert-Schmidt Norms. http://www. gatsby.ucl.ac.uk/~gretton/papers/GreBouSmoSch05.pdf

[7] Flaxman, S., Sejdinovic, D., Cunningham. J.P., Filippi, S. (2016) Bayesian Learning of Kernel Embeddings. https://stat.columbia. edu/~cunningham/pdf/FlaxmanUAI2016.pdf

[8] Greenfeld, D., Shalit, U. (2020) Robust Learning with the Hilbert-Schmidt Independence Criterion. https://arxiv.org/pdf/1910.00270.pdf

[9] Garreau, D. and Jitkrittum, W. and Kanagawa, M. (2018) Large sample analysis of the median heuristic. https://arxiv.org/pdf/1707.07269.pdf

[10] Wan-Duo Kurt Ma, J.P. Lewis, W. Bastiaan Kleijn. (2019) The HSIC Bottleneck: Deep Learning without Back-Propagation. https://arxiv.org/pdf/1908.01580v1.pdf