
Learning Interpretable Characteristic Kernels via Decision Forests

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Decision forests are popular tools for classification and regression. These forests
2 naturally generate proximity matrices that measure the frequency of observations
3 appearing in the same leaf node. While other kernels are known to have strong the-
4 oretical properties such as being characteristic, there is no similar result available
5 for decision forest-based kernels. In addition, existing approaches to independence
6 and k-sample testing may require unfeasibly large sample sizes and are not inter-
7 pretable. In this manuscript, we prove that the decision forest induced proximity
8 is a characteristic kernel, enabling consistent independence and k-sample testing
9 via decision forests. We leverage this to introduce kernel mean embedding random
10 forest (KMERF), which is a valid and consistent method for independence and
11 k-sample testing. Our extensive simulations demonstrate that KMERF outperforms
12 other tests across a variety of independence and two-sample testing scenarios.
13 Additionally, the test is interpretable, and its key features are readily discernible.
14 This work therefore demonstrates the existence of a test that is both more powerful
15 and more interpretable than existing methods, flying in the face of conventional
16 wisdom of the trade-off between the two.

17 1 Introduction

18 Decision forests are ensemble method popularized by Breiman [3]. It is highly effective in classifica-
19 tion and regression tasks, particularly in high-dimensional settings [5, 6, 37]. This is achieved by
20 randomly partitioning the feature set and using subsampling techniques to construct multiple decision
21 trees from the training data. To measure the similarity between two observations, a proximity matrix
22 can be constructed, defined as the percentage of trees in which both observations lie in the same leaf
23 node [4]. This proximity matrix serves as an induced kernel or similarity matrix for the decision
24 forest. In general, any random partition algorithm may produce such a kernel matrix.

25 As the complexity of datasets grow, it becomes increasingly necessary to develop methods that can
26 efficiently perform independence and k-sample testing. We also desire methods that are interpretable,
27 lending insight into how and why statistically significant results were determined. Parametric methods
28 are often highly interpretable, such as Pearson’s correlation and its rank variants [22, 33, 16]. These
29 methods are still popular to detect linear and monotonic relationships in univariate settings, but
30 they are not consistent for detecting more complicated nonlinear relationships. Nonparametric
31 methods can be very powerful. The more recent distance correlation (Dcorr) [36, 35] and the
32 kernel correlation (HSIC) [10, 11] are consistent for testing independence against any distribution of
33 finite second moments for any finite dimensionality; moreover, the energy-based statistics (such as
34 Dcorr) and kernel-based statistics (such as HSIC) are known to be exactly equivalent for all finite
35 samples [21, 30]. The theory supporting universal consistency of these methods (which we refer to as
36 kernel methods hereafter, without loss of generality) depends on those kernels being characteristic

kernel [28, 18, 19, 30]. Unfortunately, the above tests do not attempt to further characterize the dependency structure. To the best of our knowledge, very few tests exist [38, 15].

In addition, although these methods all have asymptotic guarantees, for finite samples, performance can be impaired by poorly choosing a particular characteristic kernel. Choosing an appropriate kernel that properly summarize geometries within the data is often times non-obvious [27]. High-dimensional data is particularly vexing [25, 38], and a number of extensions have been proposed to achieve better power such as adaptive metric kernel choice [12], low-dimensional projections [14], and marginal correlations [29].

In this paper, we leverage the popular random forest method [3] and a recent chi-square test [32] for a more powerful and interpretable method for hypothesis testing. We prove that the random forest induced kernel is a characteristic kernel, and the resulting kernel mean embedding random forest (KMERF) is a valid and consistent method for independence and k-sample testing. We then demonstrate its empirical advantage over existing tools for high-dimensional testing in a variety of dependence settings, suggesting that it will often be more powerful than existing approaches in real data. As random forest can directly estimate feature importances [3], the outputs of KMERF are also interpretable, KMERF therefore flies in the face of conventional wisdom that one must choose between power and interpretability: KMERF is both empirically more powerful and more interpretable than existing approaches.

2 Preliminaries

2.1 Hypothesis Testing

The testing independence hypothesis is formulated as follows: suppose $x_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}^q$, and n samples of $(x_i, y_i) \stackrel{iid}{\sim} F_{XY}$, i.e., x_i and y_i are realizations of random variables X and Y . The hypothesis for testing independence is

$$\begin{aligned} H_0 : F_{XY} &= F_X F_Y, \\ H_A : F_{XY} &\neq F_X F_Y. \end{aligned}$$

Given any kernel function $k(\cdot, \cdot)$, we can formulate the kernel induced correlation measure as $c_k^n(\mathbf{x}, \mathbf{y})$ using the sample kernel matrices [10, 30], where $\mathbf{x} = \{x_i\}$ and $\mathbf{y} = \{y_i\}$. When the kernel function $k(\cdot, \cdot)$ is characteristic, it has been shown that $c_k^n(\mathbf{x}, \mathbf{y}) \rightarrow 0$ if and only if \mathbf{x} and \mathbf{y} are independent [10].

The k-sample hypothesis is formulated as follows: let $w_i^j \in \mathbb{R}^p$ be the realization of random variable U_j for $j = 1, \dots, l$ and $i = 1, \dots, n_j$. Suppose the l datasets that are sampled i.i.d. from F_1, \dots, F_l and independently from one another. Then,

$$\begin{aligned} H_0 : F_1 &= F_2 = \dots = F_l, \\ H_A : \exists j \neq j' \text{ s.t. } &F_j \neq F_{j'}. \end{aligned}$$

By concatenating the l datasets and introducing an auxiliary random variable, the kernel correlation measure can be used for k-sample testing [21].

2.2 Characteristic Kernel

Definition 1. Let \mathcal{X} be a separable metric space, such as \mathbb{R}^p . A kernel function $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ measures the similarity between two observations in \mathcal{X} , and an $n \times n$ kernel matrix for $\{x_i \in \mathcal{X}, i = 1, \dots, n\}$ is defined by $\mathbf{K}(i, j) = k(x_i, x_j)$.

- A kernel $k(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is positive definite if, for any $n \geq 2$, $x_1, \dots, x_n \in \mathcal{X}$ and $a_1, \dots, a_n \in \mathbb{R}$, it satisfies

$$\sum_{i,j=1}^n a_i a_j k(x_i, x_j) \geq 0.$$

- A characteristic kernel is a positive definite kernel that has the following property: for any two random variables X_1 and X_2 with distributions F_{X_1} and F_{X_2} ,

$$E[k(\cdot, X_1)] = E[k(\cdot, X_2)] \text{ if and only if } F_{X_1} = F_{X_2}. \quad (1)$$

77 3 KMERF

78 The proposed approach for hypothesis testing, KMERF, involves the following steps:

- 79 1. Run random forest with m trees, with independent bootstrap samples of size $n_b \leq n$ used
80 to construct each tree. The tree structures (partitions) within the forest \mathbf{P} are denoted as
81 $\phi_w \in \mathbf{P}$, where $w \in 1, \dots, m$ and $\phi_w(x_i)$ represents the partition assigned to x_i .
- 82 2. Calculate the proximity kernel by

$$\mathbf{K}_{ij}^{\mathbf{x}} = \frac{1}{m} \sum_{w=1}^m [\mathbf{I}(\phi_w(x_i) = \phi_w(x_j))],$$

83 where $\mathbf{I}(\cdot)$ is the indicator function that checks whether the two observations lie in the same
84 partition in each tree.

- 85 3. Compute the unbiased kernel transformation [34, 32] on $\mathbf{K}^{\mathbf{x}}$. Namely, let

$$\mathbf{L}_{ij}^{\mathbf{x}} = \begin{cases} \mathbf{K}_{ij}^{\mathbf{x}} - \frac{1}{n-2} \sum_{t=1}^n \mathbf{K}_{it}^{\mathbf{x}} - \frac{1}{n-2} \sum_{s=1}^n \mathbf{K}_{sj}^{\mathbf{x}} + \frac{1}{(n-1)(n-2)} \sum_{s,t=1}^n \mathbf{K}_{st}^{\mathbf{x}} & i \neq j \\ 0 & i = j \end{cases}$$

- 86 4. Let $\mathbf{K}^{\mathbf{y}}$ be the Euclidean distance induced kernel by Shen and Vogelstein [30], or the
87 proximity kernel in the case that dimensions of \mathbf{x} and \mathbf{y} is the same, that is $p = q$, and
88 compute $\mathbf{L}^{\mathbf{y}}$ using the same unbiased transformation. Then the KMERF statistic for the
89 induced kernel k is,

$$c_k^n(\mathbf{x}, \mathbf{y}) = \frac{1}{n(n-3)} \text{trace}(\mathbf{L}^{\mathbf{x}} \mathbf{L}^{\mathbf{y}}).$$

- 90 5. Compute the p-value via the following chi-square test [32]:

$$p = 1 - F_{\chi_1^2-1} \left(n \cdot \frac{c_k^n(\mathbf{x}, \mathbf{y})}{\sqrt{c_k^n(\mathbf{x}, \mathbf{x}) \cdot c_k^n(\mathbf{y}, \mathbf{y})}} \right),$$

91 where χ_1^2 is the chi-square distribution of degree 1. Reject the independence hypothesis if
92 the p-value is less than a specified type 1 error level, say 0.05.

93 In the numerical implementation, the standard supervised random forest is used with $m = 500$ (which
94 is also applicable to the unsupervised version or other random forest variants [2, 1, 37]). In the second
95 step, we simply compute the proximity kernel defined by the random forest induced kernel. In the
96 third step, we normalize the proximity kernel to ensure it obtains a consistent dependence measure;
97 this is the KMERF test statistic. We found that utilizing the multiscale version of the kernel correlation
98 [38, 31], which is equivalent for linear relationships while being better for nonlinear relationships,
99 produced similar results to using distance correlation, but substantially increased runtimes.

100 Note that one could also compute a p-value for KMERF via the permutation test, which is a standard
101 procedure for testing independence [9]. Specifically, first compute a kernel on the observed $\{x_i\}$ and
102 $\{y_i\}$. Then randomly permute the index of $\{y_i\}$, repeat the kernel generation process for $\{y_i\}$ for
103 each permutation. This process involves training a new random forest for each permutation. Finally,
104 compute the test statistic for each of the permutations, and the p-value equals the percentage the
105 permuted statistics that are larger than the observed statistic. However, the permutation test is very
106 slow for large sample size and almost always yields similar results as the chi-square test.

107 4 Theoretical Properties

108 Here, we show that the random forest kernel characteristic, and the induced test statistic used in
109 KMERF allows for valid and universally consistent independence and k-sample testing. All proofs
110 are in appendix.

111 For a kernel to be characteristic, it first needs to be positive definite, which is indeed the case for the
112 forest-induced kernel:

113 **Theorem 1.** *The random forest induced kernel \mathbf{K}^x is always positive definite.*

114 This theorem holds because the forest-induced kernel is a summation of a permuted block diagonal
 115 matrix, with each matrix coming from individual tree, that is positive definite [7]; and a summation
 116 of positive definite matrices is still positive definite.

117 Next, we show the kernel is characteristic when the tree partition area converges to zero. A similar
 118 property is also used for proving classification consistency for k-nearest-neighbors [8], and we shall
 119 denote $N(\phi_w)$ as the maximum area of each part.

120 **Theorem 2.** *Suppose as $n, m \rightarrow \infty$, $N(\phi_w) \rightarrow 0$ for each tree $\phi_w \in \mathbf{P}$ and each observation x_i .
 121 Then the random forest induced kernel \mathbf{K}^x is asymptotically characteristic.*

122 Intuitively, for sufficiently many trees and sufficiently small leaf region, observations generated by
 123 two different distributions cannot always be in the same leaf region.

124 This leads to the validity and consistency result of KMERF:

125 **Corollary 2.1.** *KMERF satisfies*

$$\lim_{n \rightarrow \infty} c_k^n(\mathbf{x}, \mathbf{y}) = c \geq 0,$$

126 *with equality to 0 if and only if $F_{XY} = F_X F_Y$. Moreover, for sufficiently large n and sufficiently
 127 small type 1 error level α , this method is valid and consistent for independence and k -sample testing.*

128 By Gretton et al. [10], any characteristic-kernel based dependence measure converges to 0 if and
 129 only if X and Y are independent. Moreover, Shen et al. [32] showed that the chi-square distribution
 130 $\chi_1^2 - 1$ approximates and upper-tail dominates the true null distribution of any unbiased kernel when
 131 using distance correlation, making it a valid and consistent test.

132 5 Simulations

133 In this section we exhibit the consistency and validity of KMERF, and compare its testing power
 134 with other competitors in a comprehensive simulation set-up. We utilize the `hyppo` package in
 135 Python [20], which uses `scikit-learn` [23] random forest with 500 trees and otherwise default
 136 hyper-parameters, and calculate the proximity matrix from this. The KMERF statistic and p-value
 137 then computed via the process in Section 3. The mathematical details for each simulation type is in
 138 the Appendix C.

139 5.1 Testing Independence

140 In this section we compare KMERF to Multiscale Graph Correlation (MGC), Distance Correlation
 141 (Dcorr), Hilbert-Schmidt Independence Criterion (Hsic), and Heller-Heller-Gorfine (HHG) method,
 142 Canonical Correlation Analysis (CCA), and the RV coefficient. The HHG method has been shown
 143 to work extremely well against nonlinear dependencies [13]. The MGC method has been shown
 144 to work well against linear, nonlinear, and high-dimensional dependencies [31]. The CCA and RV
 145 coefficients are popular multivariate extensions of Pearson correlation. For each method, we use the
 146 corresponding implementation in `hyppo` with default settings.

147 We take 20 high-dimensional simulation settings [38], consisting of various linear, monotone, and
 148 strongly nonlinear dependencies with p increasing, $q = 1$, and $n = 100$. To estimate the testing
 149 power in each setting, we generate dependent (x_i, y_i) for $i = 1, \dots, n$, compute the test statistic
 150 for each method, repeat for $r = 10000$ times. Via the empirical alternative and null distribution of
 151 the test statistic, we estimate the testing power of each method at type 1 error level of $\alpha = 0.05$.
 152 The power result is shown in Figure 1 shows that KMERF achieves superior performance for most
 153 simulation modalities, except a few like circle and ellipse.

154 5.2 Two Sample Testing

155 Here, we compare the performance in the two-sample testing regime. It has been shown that
 156 all independence measures can be used for two-sample testing [21, 30], allowing all previous
 157 independence testing methods to be compared here as well. Once again, we investigate statistical

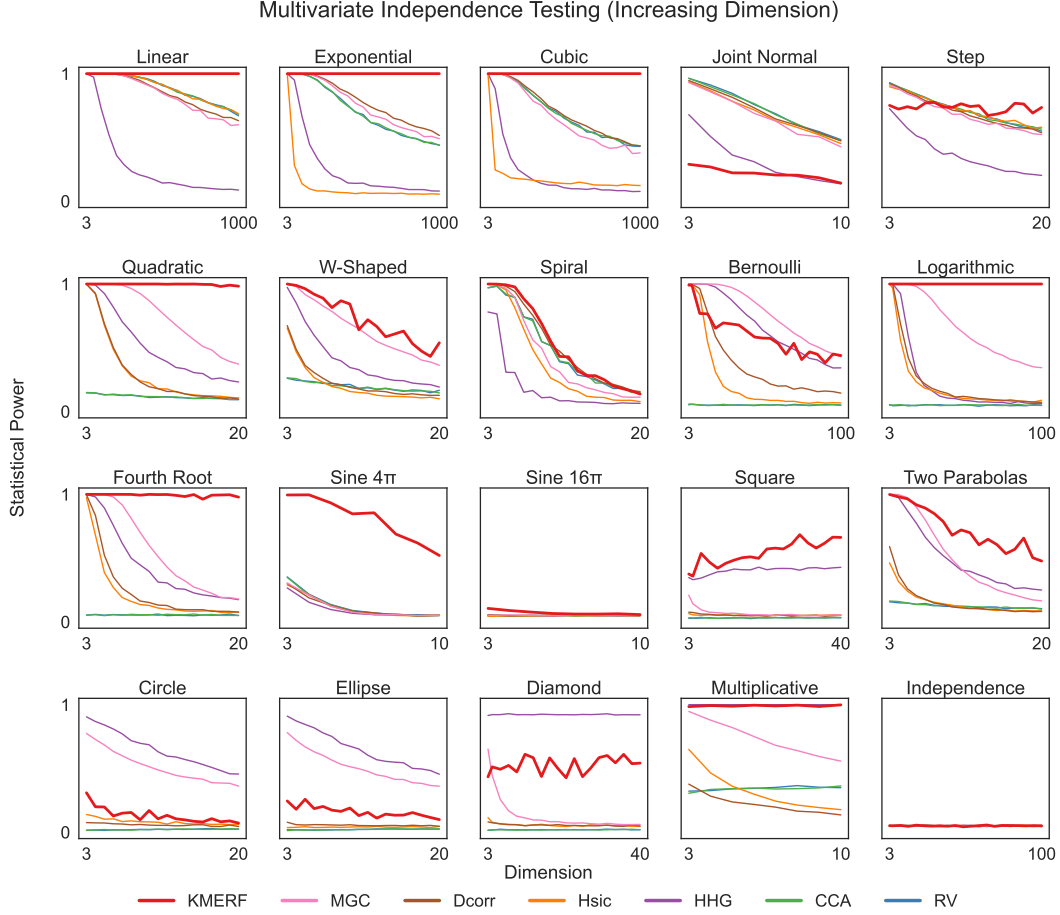


Figure 1: Multivariate independence testing power for 20 different settings with increasing p , fixed $q = 1$, and $n = 100$. For the majority of the simulations and simulation dimensions, KMERF performs as well as, or better than, existing multivariate independence tests in high-dimensional dependence testing.

power differences with 20 simulation settings consisting of various linear and nonlinear, monotonic and nonmonotonic functions with dimension increasing from $p = 3, \dots, 10$, $q = 1$, and $n = 100$. We then apply a random rotation to this generated simulation and generate the second independent sample (via a rigid transformation).

Figure 2 shows that, once again, for the majority of simulations settings, KMERF performs at or better than other tests in nearly all simulations and simulation dimensions. For certain simulation settings, especially the exponential, cubic, and fourth root, KMERF vastly outperforms other metrics as dimensions increases.

5.3 Interpretability

Not only does KMERF typically offer empirically better statistical power compared to alternatives, it also offers insights into which features are the most important within the data set. Figure 3 shows normalized 95% confidence intervals of relative feature importances for each simulation, where the black line shows the mean and the light grey line shows the 95% confidence interval. Mean and individual tree feature importances were normalized using min-max feature scaling. The simulations were modified such that the weighting of each feature decreased as feature importance increased, with the expectation that the algorithm would detect a decrease in feature importance as dimension increased. With these simulations, we are able to determine that exact feature importance trend,

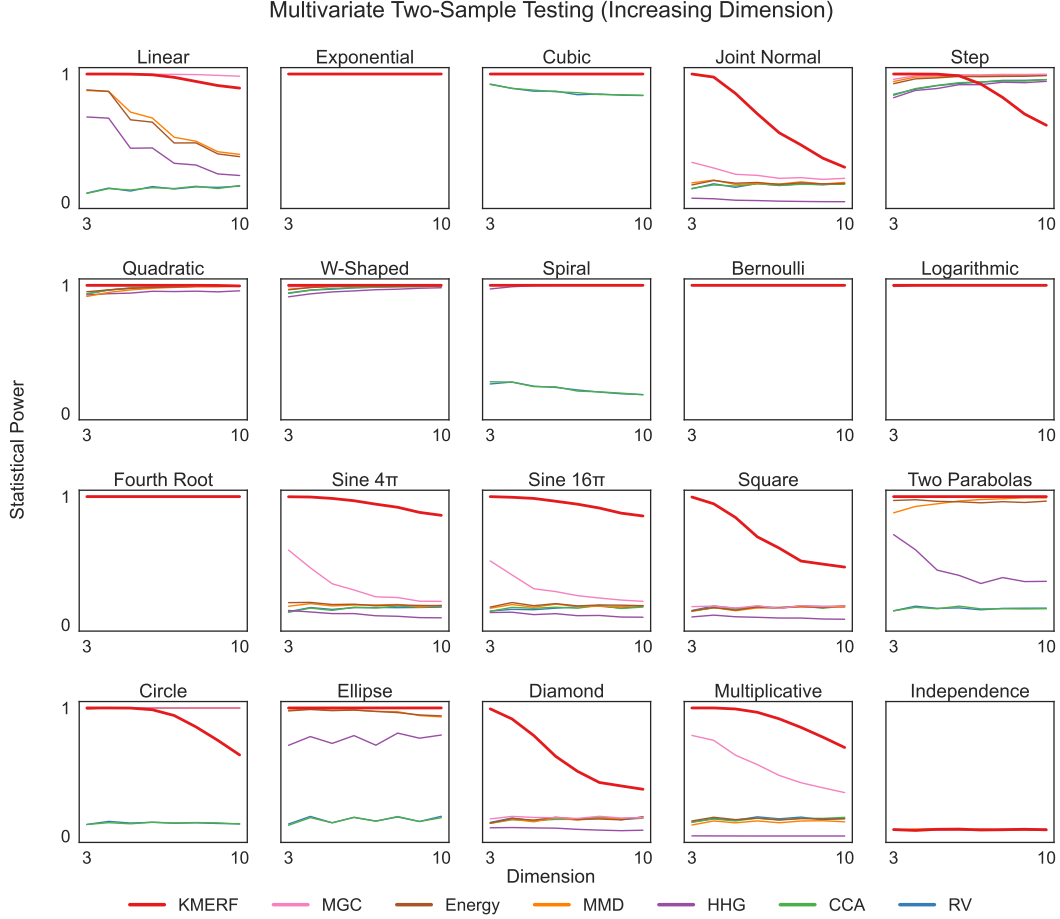


Figure 2: Multivariate two-sample testing power for 20 different settings with increasing p , fixed $q = 1$, and $n = 100$. For nearly all simulations and simulation dimensions, KMERF performs as well as, or better than, existing multivariate two-sample tests in high-dimensional dependence testing.

except for a few of the more complex simulations. The process we used to generate this figure can be trivially extended to a two-sample or k-sample case.

6 Real Data

We then applied KMERF to a data set consisting of proteolytic peptides derived from the blood samples of 95 individuals harboring pancreatic ($n=10$), ovarian ($n=24$), colorectal cancer ($n=28$), and healthy controls ($n=33$) [38]. The processed data included 318 peptides derived from 121 proteins (see Appendix D for full details). Figure 4 shows the p-values for KMERF between pancreatic and healthy subjects compared to the p-values for KMERF between pancreatic cancer and all other subjects. The test identifies neurogranin as a potentially valuable marker for pancreatic cancer, which the literature also corroborates [41, 40]. Meanwhile, while some of the other tests identified this biomarker, they identified others that are upregulated in other types of cancers as well (false positives). We also show in the figure that the biomarker chosen by KMERF provides better true positive detection when compared to the other tests (there is no ground truth in this case, so a leave-one-out k-nearest-neighbor classification approach was used instead).

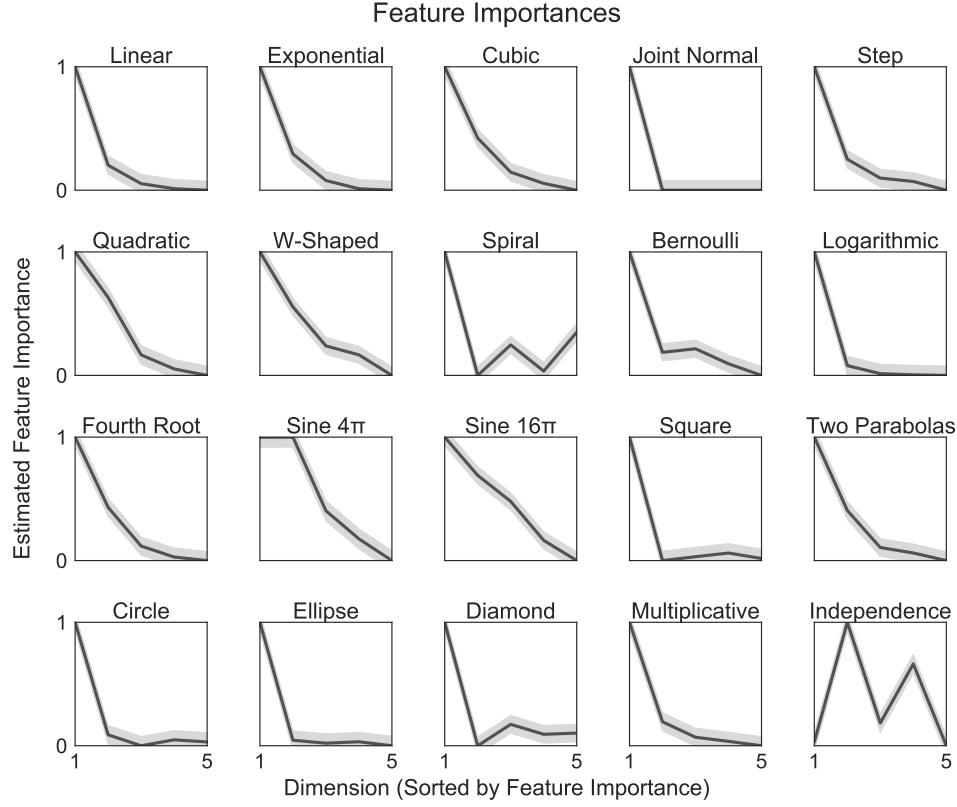


Figure 3: Normalized mean (black) and 95% confidence intervals (light grey) using min-max normalization for relative feature importances derived from random forest over five dimensions for each simulation tested for 100 samples. The features were sorted from most to least informative for all simulations except for the Independence simulation). As expected, estimated feature importance decreases as dimension increases. A feature of KMERF is insights into interpretability, and we show here which dimensions of our simulations influence the outcome of independence test the most.

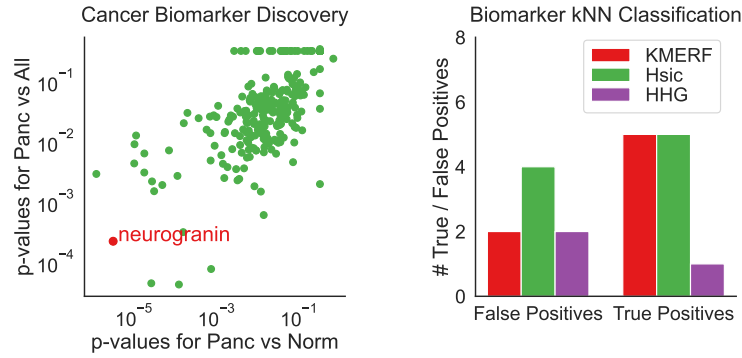


Figure 4: (A) For each peptide, the p-values for testing dependence between pancreatic and healthy subjects by KMERF is compared to the p-value for testing dependence between pancreatic and all other subjects. At the critical level 0.05, KMERF identifies a unique protein. (B) The true and false positive counts using a k-nearest neighbor (choosing the best $k \in [1, 10]$) leave-one-out classification using only the significant peptides identified by each method. The peptide identified by KMERF achieves the best true and false positive rates.

7 Discussion

KMERF is, to the best of our knowledge, one of the first learned kernel that is proven to be characteristic. The empirical experiments presented here illustrate the potential advantages of learning kernels, specifically for independence and k-sample testing.

In fact, multiscale graph correlation [38, 31] can be thought of, in a sense, as kernel learning: given n samples, and a pair of kernel or distance functions, it chooses one of the approximately n^2 sparsified kernels, by excluding all but the nearest neighbors for each data point [38, 31]. Because random forest can be thought of as a nearest neighbor algorithm [17], in a sense, the forest induced kernel is a natural extension of Vogelstein et al. [38], which leads to far more data-adaptive estimates of the nearest neighbors using supervised information. Moreover, proving that the random-forest induced kernel is characteristic is a first step towards building lifelong learning kernel machines with strong theoretical guarantees [24, 39].

As the choice of kernel is crucial for empirical performance, this manuscript offers a new kernel construction that is not only universally consistent for testing independence, but also exhibits strong empirical advantages, especially for high-dimensional testing. What is unique to this choice of kernel is the robustness and interpretability. It will be worthwhile to further understand the underlying theoretical mechanism of the induced characteristic kernel, as well as evaluating the performance of these forest induced kernels on other learning problems, including classification, regression, clustering, and embedding [26].

References

- [1] S. Athey, J. Tibshirani, and S. Wager. Generalized random forests. *Annals of Statistics*, 47(2): 1148–1178, 2018.
- [2] R. Blaser and P. Fryzlewicz. Random rotation ensembles. *Journal of Machine Learning Research*, 17(4):1–26, 2016.
- [3] L. Breiman. Random forests. *Machine Learning*, 4(1):5–32, October 2001.
- [4] L. Breiman. Some infinity theory for predictor ensembles. *Journal of Combinatorial Theory, Series A*, 98:175–191, 2002.
- [5] R. Caruana and A. Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, pages 161–168. ACM, 2006.
- [6] R. Caruana, N. Karampatziakis, and A. Yessenalina. An empirical evaluation of supervised learning in high dimensions. *Proceedings of the 25th International Conference on Machine Learning*, 2008.
- [7] A. Davies and Z. Ghahramani. The random forest kernel and creating other kernels for big data from random partitions. *arXiv:1402.4293v1*, 2014.
- [8] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer, 1996.
- [9] P. Good. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. Springer, 2005.
- [10] A. Gretton, R. Herbrich, A. Smola, O. Bousquet, and B. Schölkopf. Kernel methods for measuring independence. *Journal of Machine Learning Research*, 6:2075–2129, 2005.
- [11] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A Kernel Two-Sample Test. *Journal of Machine Learning Research*, 13(25):723–773, 2012. ISSN 1533-7928.
- [12] A. Gretton, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. M. Pontil, K. Fukumizu, and B. K. Sriperumbudur. Optimal kernel choice for large-scale two-sample tests. In *Advances in neural information processing systems 25*, pages 1205–1213, 2012.
- [13] R. Heller, Y. Heller, and M. Gorfine. A consistent multivariate test of association based on ranks of distances. *Biometrika*, 100(2):503–510, 2013.

- [14] C. Huang and X. Huo. A statistically and numerically efficient independence test based on random projections and distance covariance. *arXiv*, 2017.
- [15] W. Jitkrittum, Z. Szabó, K. P. Chwialkowski, and A. Gretton. Interpretable distribution features with maximum testing power. In *Advances in Neural Information Processing Systems*, pages 181–189, 2016.
- [16] M. G. Kendall. *Rank Correlation Methods*. London: Griffin, 1970.
- [17] Y. Lin and Y. Jeon. Random forests and adaptive nearest neighbors. *J. Am. Stat. Assoc.*, 101(474):578–590, 2006.
- [18] R. Lyons. Distance covariance in metric spaces. *Annals of Probability*, 41(5):3284–3305, 2013.
- [19] R. Lyons. Errata to “distance covariance in metric spaces”. *Annals of Probability*, 46(4):2400–2405, 2018.
- [20] S. Panda, S. Palaniappan, J. Xiong, E. W. Bridgeford, R. Mehta, C. Shen, and J. T. Vogelstein. hyppo: A comprehensive multivariate hypothesis testing python package, 2020.
- [21] S. Panda, C. Shen, R. Perry, J. Zorn, A. Lutz, C. E. Priebe, and J. T. Vogelstein. Nonpar manova via independence testing, 2021.
- [22] K. Pearson. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011. URL <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [24] A. Pentina and S. Ben-David. Multi-task and lifelong learning of kernels. In *Algorithmic Learning Theory*, pages 194–208. Springer International Publishing, 2015.
- [25] A. Ramdas, S. J. Reddi, B. Póczos, A. Singh, and L. Wasserman. On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions. In *29th AAAI Conference on Artificial Intelligence*, 2015.
- [26] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [27] D. Sejdinovic, A. Gretton, B. Sriperumbudur, and K. Fukumizu. Hypothesis testing using pairwise distances and associated kernels (with appendix). *arXiv preprint arXiv:1205.0411*, 2012.
- [28] D. Sejdinovic, B. Sriperumbudur, A. Gretton, and K. Fukumizu. Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *Annals of Statistics*, 41(5):2263–2291, 2013.
- [29] C. Shen. Scaling up independence testing: Maximum marginal distance correlation and the chi-square test. <https://arxiv.org/abs/2001.01095>, 2023.
- [30] C. Shen and J. T. Vogelstein. The exact equivalence of distance and kernel methods in hypothesis testing. *ASTA Advances in Statistical Analysis*, 105(3):385–403, 2021.
- [31] C. Shen, C. E. Priebe, and J. T. Vogelstein. From distance correlation to multiscale graph correlation. *Journal of the American Statistical Association*, 115(529):280–291, 2020.
- [32] C. Shen, S. Panda, and J. T. Vogelstein. The chi-square test of distance correlation. *Journal of Computational and Graphical Statistics*, 31(1):254–262, 2022.
- [33] C. Spearman. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology*, 15(1):72, 1904. ISSN 00029556. doi: 10.2307/1412159. URL <http://www.jstor.org/stable/1412159?origin=crossref>.

- 281 [34] G. Székely and M. Rizzo. Partial distance correlation with methods for dissimilarities. *Annals*
282 *of Statistics*, 42(6):2382–2412, 2014.
- 283 [35] G. Székely and M. L. Rizzo. The distance correlation t-test of independence in high dimension.
284 *Journal of Multivariate Analysis*, 117:193–213, 2013.
- 285 [36] G. Székely, M. L. Rizzo, and N. K. Bakirov. Measuring and testing independence by correlation
286 of distances. *Annals of Statistics*, 35(6):2769–2794, 2007.
- 287 [37] T. Tomita, J. Browne, C. Shen, J. Chung, J. Patsolic, B. Falk, J. Yim, C. E. Priebe, R. Burns,
288 M. Maggioni, and J. T. Vogelstein. Sparse projection oblique randomer forests. *Journal of*
289 *Machine Learning Research*, 21(104):1–39, 2020.
- 290 [38] J. T. Vogelstein, Q. Wang, E. W. Bridgeford, C. E. Priebe, M. Maggioni, and C. Shen. Discover-
291 ing and deciphering relationships across disparate data modalities. *eLife*, 8:e41690, 2019.
- 292 [39] J. T. Vogelstein, J. Dey, H. S. Helm, W. LeVine, R. D. Mehta, T. M. Tomita, H. Xu, A. Geisa,
293 Q. Wang, G. M. van de Ven, C. Gao, W. Yang, B. Tower, J. Larson, C. M. White, and C. E.
294 Priebe. Representation ensembling for synergistic lifelong learning with quasilinear complexity,
295 2023.
- 296 [40] E. A. Willemse, A. De Vos, E. M. Herries, U. Andreasson, S. Engelborghs, W. M. Van Der Flier,
297 P. Scheltens, D. Crimmins, J. H. Ladenson, E. Vanmechelen, et al. Neurogranin as cerebrospinal
298 fluid biomarker for alzheimer disease: an assay comparison study. *Clinical chemistry*, 64(6):
299 927–937, 2018.
- 300 [41] J. Yang, F. K. Korley, M. Dai, and A. D. Everett. Serum neurogranin measurement as a
301 biomarker of acute traumatic brain injury. *Clinical biochemistry*, 48(13-14):843–848, 2015.