A Semi-Supervised Kernel Two-Sample Test

Gyumin Lee[1)], Shubhanshu Shekhar[2)], and Ilmun Kim[1),3)]

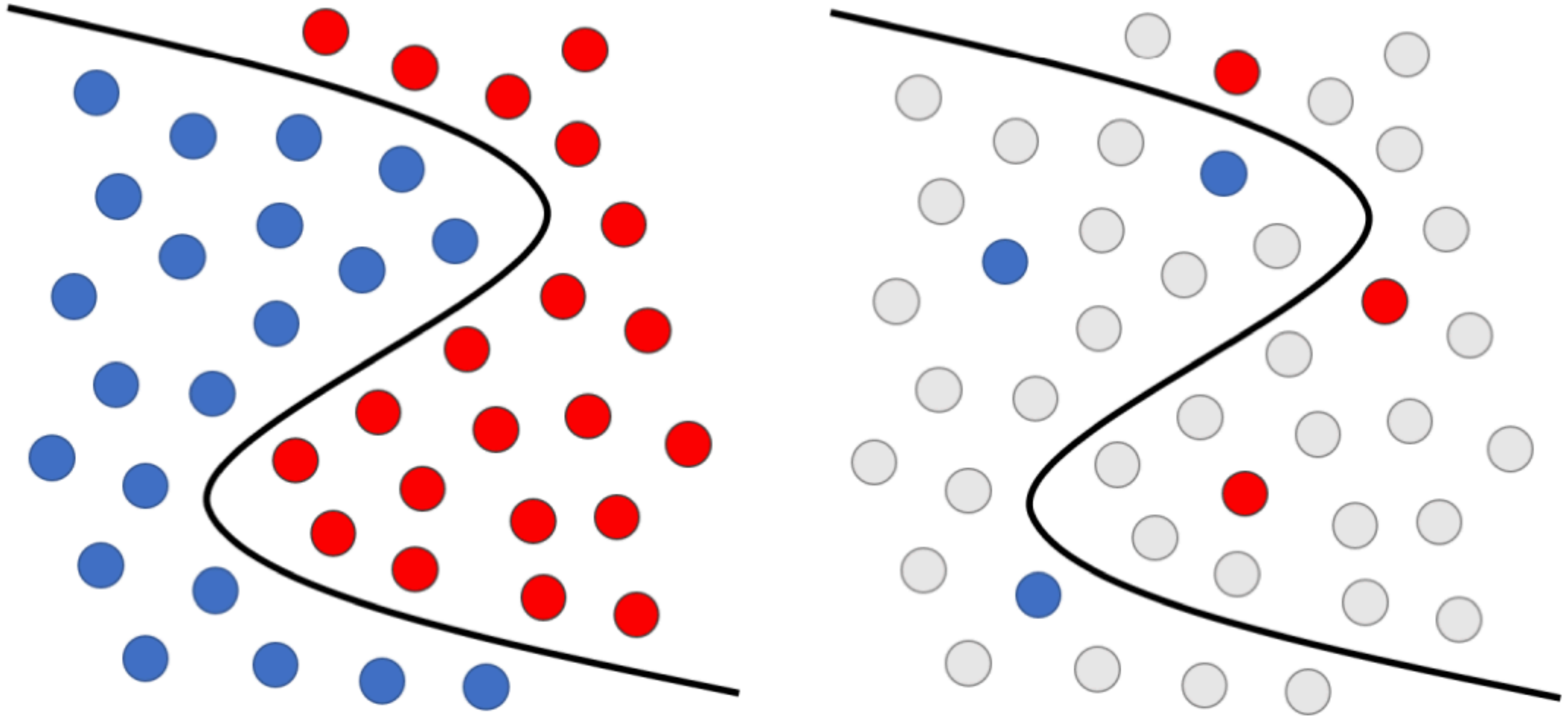
1) Department of Statistics and Data Science, Yonsei University, Seoul, South Korea.
2) Department of EECS, University of Michigan, Ann Arbor, USA.
3) Department of Applied Statistics, Yonsei University, Seoul, South Korea.

Contents

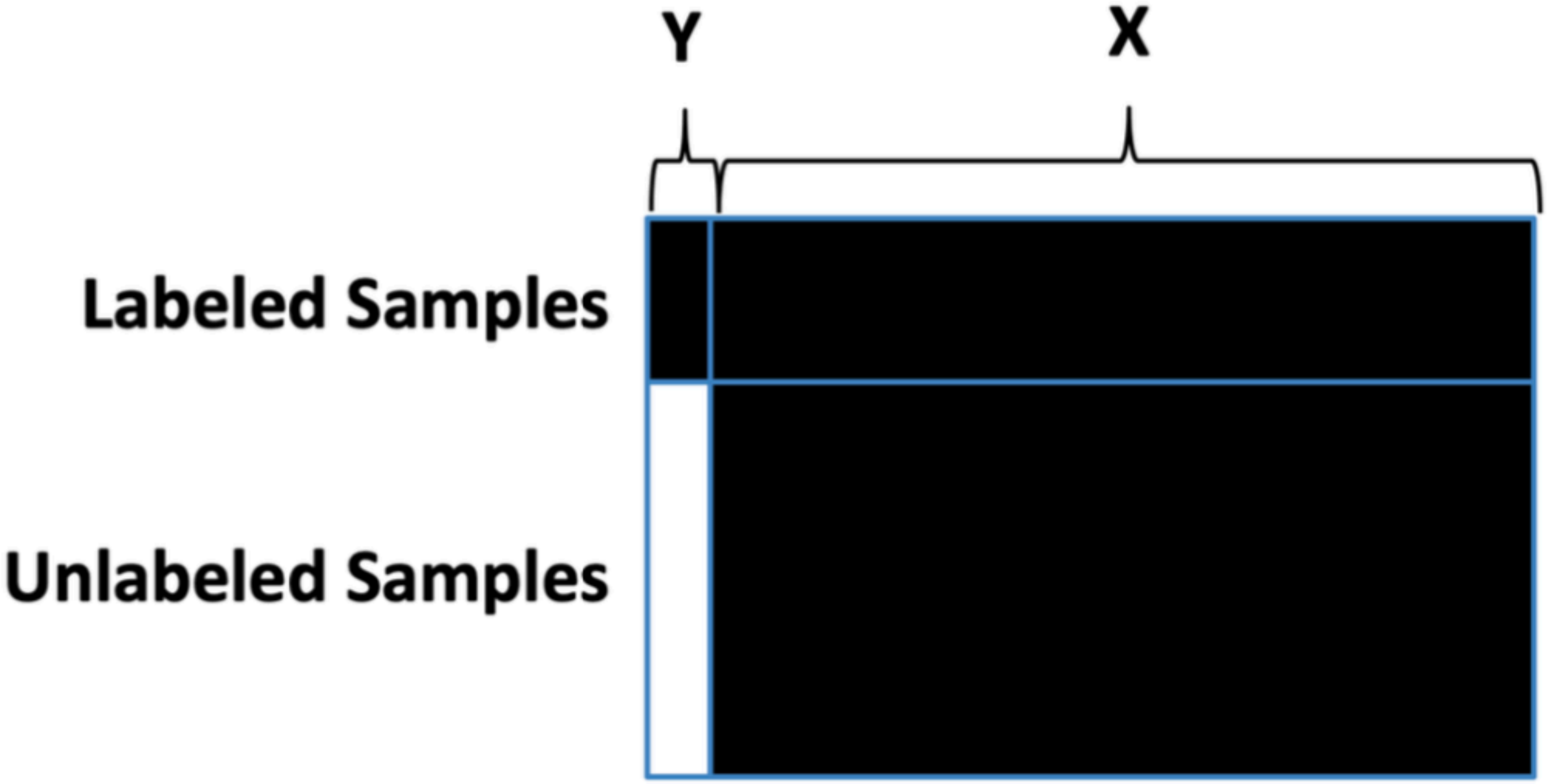
Introduction

Semi-Supervised Inference

: inference based on a (small) labeled dataset together with a (large) unlabeled dataset.



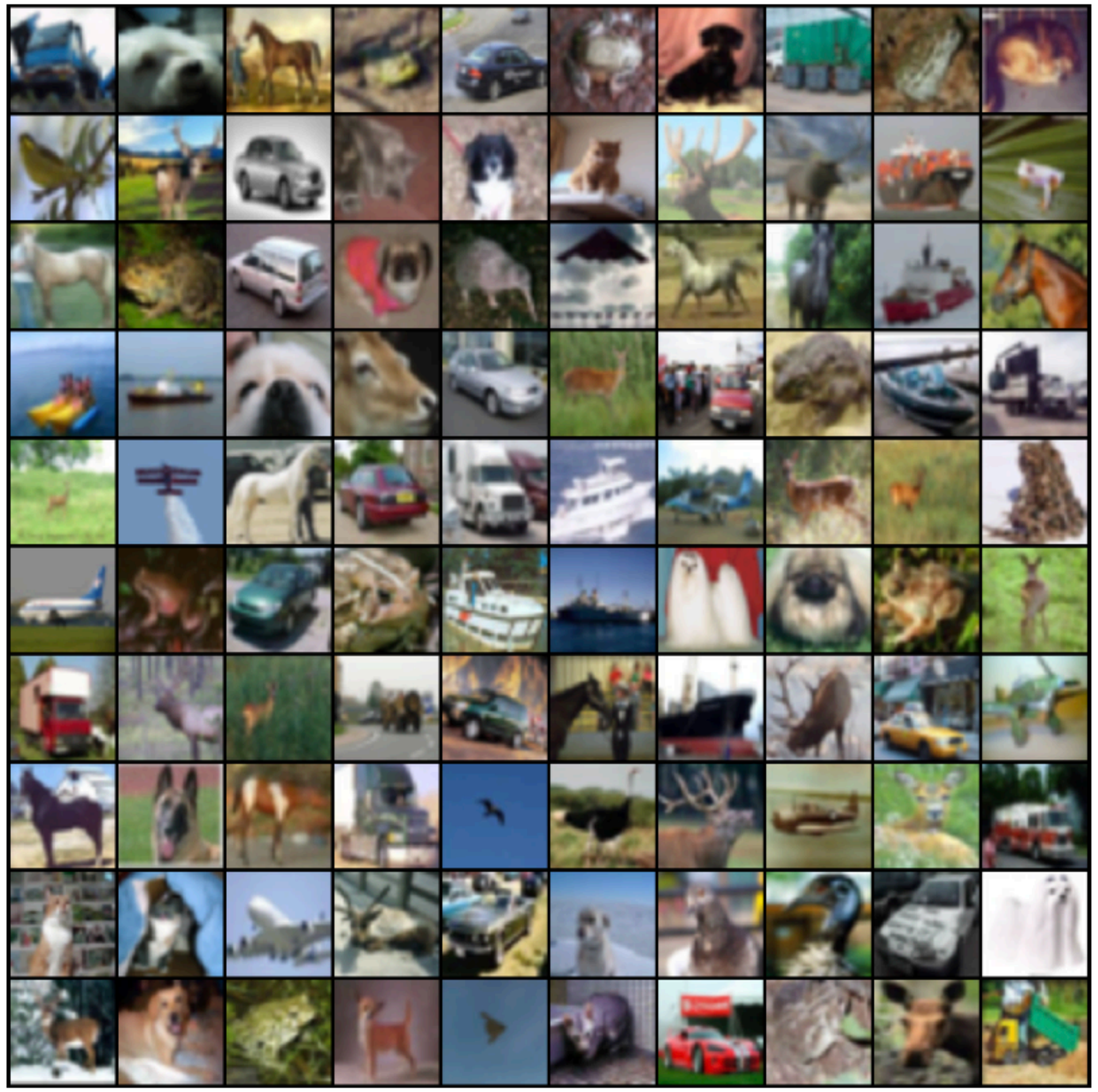
(a) Supervised Learning (b) Semi-Supervised Learning



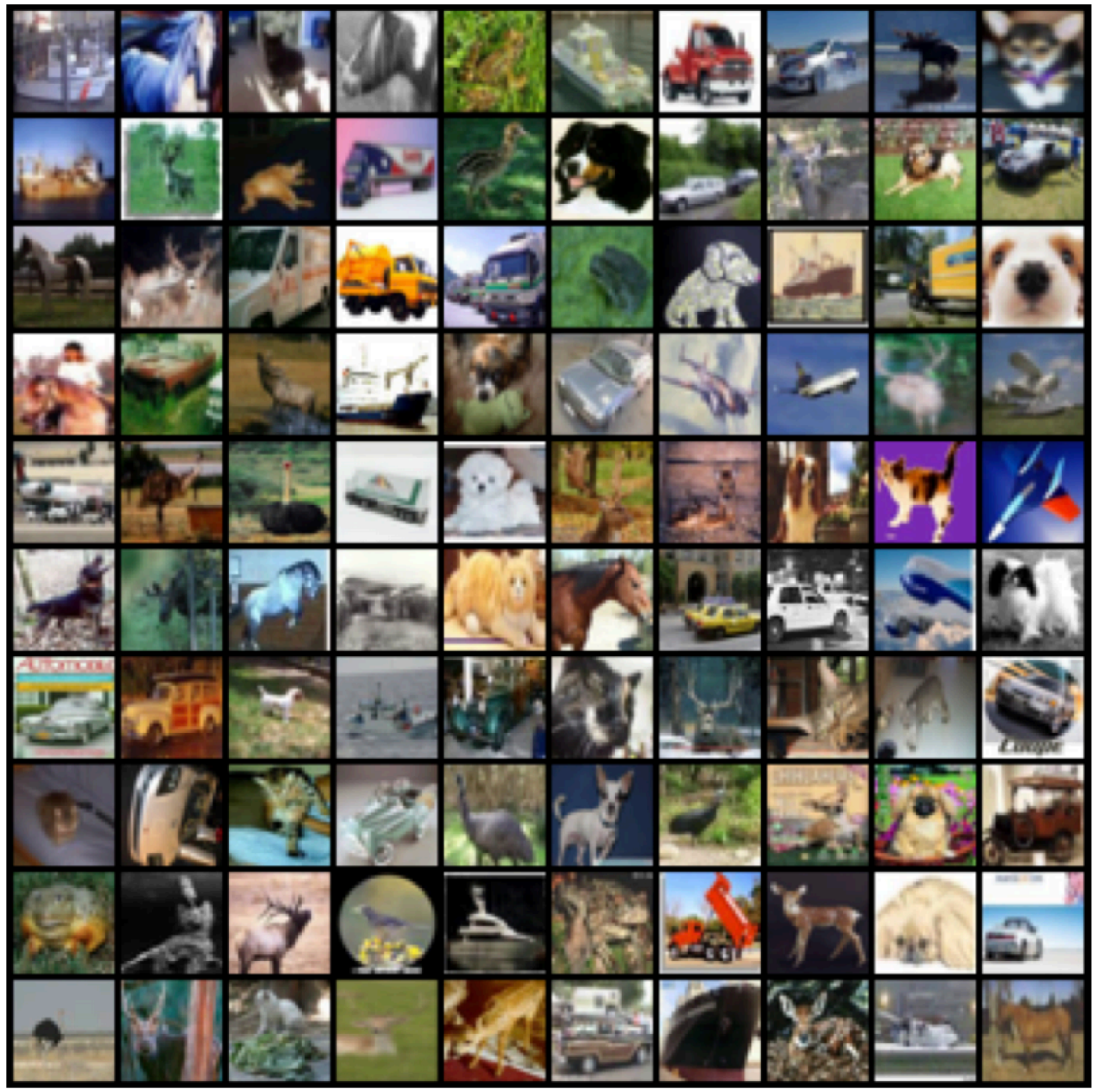
- Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R., & Jin, C. (2023). Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Computers in Biology and Medicine*, 107840.
- Zhang, A., Brown, L. D., & Cai, T. T. (2019). Semi-supervised inference: General theory and estimation of means.

Introduction

Two-sample test: $H_0 : P_X = P_Y \quad \text{vs.} \quad H_1 : P_X \neq P_Y$



(a) *CIFAR-10* test set



(b) *CIFAR-10.1* test set

X_1: Now disturbing reports out of Newfoundland show that the fragile snow crab industry is in serious decline. First the west coast salmon, the east coast salmon and the cod, and now the snow crabs off Newfoundland.

X_2: To my pleasant surprise he responded that he had personally visited those wharves and that he had already announced money to fix them. What wharves did the minister visit in my riding and how much additional funding is he going to provide for Delaps Cove, Hampton, Port Lorne,

...

$$P_X \stackrel{?}{=} Q_Y$$

Y_1: Honourable senators, I have a question for the Leader of the Government in the Senate with regard to the support funding to farmers that has been announced. Most farmers have not received any money yet.

Y_2: On the grain transportation system we have had the Estey report and the Kroeger report. We could go on and on. Recently programs have been announced over and over by the government such as money for the disaster in agriculture on the prairies and across Canada.

...

- Liu, F., Xu, W., Lu, J., Zhang, G., Gretton, A., & Sutherland, D. J. (2020, November). Learning deep kernels for non-parametric two-sample tests. In *International conference on machine learning* (pp. 6316-6326). PMLR.
- http://www.gatsby.ucl.ac.uk/~gretton/papers/columbia23/columbia23_21

Introduction

Maximum Mean Discrepancy: $\mathrm{MMD}(P,Q;\mathcal{F}) = \sup_{\substack{f\in\mathcal{F}:\|f\|\le 1 \\ \mathcal{F}\text{ is an RKHS}}} \left[\mathbb{E}_{p(\boldsymbol{x})}[f(\boldsymbol{x})] - \mathbb{E}_{q(\boldsymbol{x}')}[f(\boldsymbol{x}')]\right]$

Mean embedding: $\mu_P = [\ldots \mathbb{E}_P[\varphi_i(X)]\ldots]$ such that $\mathbb{E}_P(f(X)) = \langle f, \mu_P\rangle_{\mathcal{F}}$

We express MMD with the distance between mean embeddings

$$\begin{aligned}\mathrm{MMD}^2(P,Q) &= \|\mu_P - \mu_Q\|^2_{\mathcal{F}} \\ &= \underbrace{\mathbb{E}_P[k(x,x')]}_{(\mathrm{a})} + \underbrace{\mathbb{E}_Q[k(y,y')]}_{(\mathrm{a})} - 2\underbrace{\mathbb{E}_{P,Q}[k(x,y)]}_{(\mathrm{b})},\end{aligned}$$

, and use an unbiased empirical estimator of MMD as

$$\mathrm{MMD}_u^2[\mathcal{F},X,Y] = \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k(x_i,x_j) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k(y_i,y_j) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i,y_j)$$

- http://www.gatsby.ucl.ac.uk/~gretton/papers/columbia23/columbia23_21
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, *13*(1), 723-773.
- Murphy, K. P. (2018). Machine learning: A probabilistic perspective (adaptive computation and machine learning series). *The MIT Press: London, UK*.

Introduction

Maximum Mean Discrepancy: $\mathrm{MMD}(P,Q;\mathcal{F}) = \sup_{\substack{f\in\mathcal{F}:\|f\|\le 1 \\ \mathcal{F} \text{ is an RKHS}}} \left[\mathbb{E}_{p(\boldsymbol{x})}[f(\boldsymbol{x})] - \mathbb{E}_{q(\boldsymbol{x}')}\left[f\left(\boldsymbol{x}'\right)\right]\right]$

Mean embedding: $\mu_P = [\ldots \mathbb{E}_P\left[\varphi_i(X)\right]\ldots]$ such that $\mathbb{E}_P(f(X)) = \langle f, \mu_P\rangle_{\mathcal{F}}$

We express MMD with the distance between mean embeddings

$$\begin{aligned}\mathrm{MMD}^2(P,Q) &= \|\mu_P - \mu_Q\|_{\mathcal{F}}^2 \\ &= \underbrace{\mathbb{E}_P[k(x,x')]}_{\text{(a)}} + \underbrace{\mathbb{E}_Q[k(y,y')]}_{\text{(a)}} - 2\underbrace{\mathbb{E}_{P,Q}[k(x,y)]}_{\text{(b)}},\end{aligned}$$

, and use an unbiased empirical estimator of MMD as

$$\mathrm{MMD}_u^2[\mathcal{F},X,Y] = \frac{1}{m(m-1)}\sum_{i=1}^{m}\sum_{j\neq i}^{m} k\left(x_i,x_j\right) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} k\left(y_i,y_j\right) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k\left(x_i,y_j\right)$$

- http://www.gatsby.ucl.ac.uk/~gretton/papers/columbia23/columbia23_21
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, *13*(1), 723-773.
- Murphy, K. P. (2018). Machine learning: A probabilistic perspective (adaptive computation and machine learning series). *The MIT Press: London, UK*.

Introduction

Maximum Mean Discrepancy: $\mathrm{MMD}(P, Q; \mathcal{F}) = \sup_{\substack{f \in \mathcal{F}: \|f\| \leq 1 \\ \mathcal{F} \text{ is an RKHS}}} \left[\mathbb{E}_{p(\boldsymbol{x})}[f(\boldsymbol{x})] - \mathbb{E}_{q(\boldsymbol{x}')}[f(\boldsymbol{x}')]\right]$

Mean embedding: $\mu_P = [\ldots \mathbb{E}_P[\varphi_i(X)] \ldots]$ such that $\mathbb{E}_P(f(X)) = \langle f, \mu_P \rangle_{\mathcal{F}}$

We express MMD with the distance between mean embeddings

$$\begin{aligned} \mathrm{MMD}^2(P, Q) &= \|\mu_P - \mu_Q\|_{\mathcal{F}}^2 \\ &= \underbrace{\mathbb{E}_P[k(x, x')]}_{\text{(a)}} + \underbrace{\mathbb{E}_Q[k(y, y')]}_{\text{(a)}} - 2\underbrace{\mathbb{E}_{P,Q}[k(x, y)]}_{\text{(b)}}, \end{aligned}$$

, and use an unbiased empirical estimator of MMD as

$$\mathrm{MMD}_u^2[\mathcal{F}, X, Y] = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j \neq i}^{m} k(x_i, x_j) + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k(y_i, y_j) - \frac{2}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} k(x_i, y_j)$$

- http://www.gatsby.ucl.ac.uk/~gretton/papers/columbia23/columbia23_21
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, *13*(1), 723-773.
- Murphy, K. P. (2018). Machine learning: A probabilistic perspective (adaptive computation and machine learning series). *The MIT Press: London, UK*.

Introduction

Maximum Mean Discrepancy: $\mathrm{MMD}(P, Q; \mathcal{F}) = \sup_{\substack{f \in \mathcal{F}: \|f\| \leq 1 \\ \mathcal{F} \text{ is an RKHS}}} \left[\mathbb{E}_{p(\boldsymbol{x})}[f(\boldsymbol{x})] - \mathbb{E}_{q(\boldsymbol{x}')}\left[f\left(\boldsymbol{x}'\right)\right]\right]$

Mean embedding: $\mu_P = [\ldots \mathbb{E}_P [\varphi_i(X)] \ldots]$ such that $\mathbb{E}_P(f(X)) = \langle f, \mu_P \rangle_{\mathcal{F}}$

We express MMD with the distance between mean embeddings

$$\begin{aligned} \mathrm{MMD}^2(P, Q) &= \|\mu_P - \mu_Q\|_{\mathcal{F}}^2 \\ &= \underbrace{\mathbb{E}_P[k(x, x')]}_{\text{(a)}} + \underbrace{\mathbb{E}_Q[k(y, y')]}_{\text{(a)}} - 2 \underbrace{\mathbb{E}_{P,Q}[k(x, y)]}_{\text{(b)}}, \end{aligned}$$

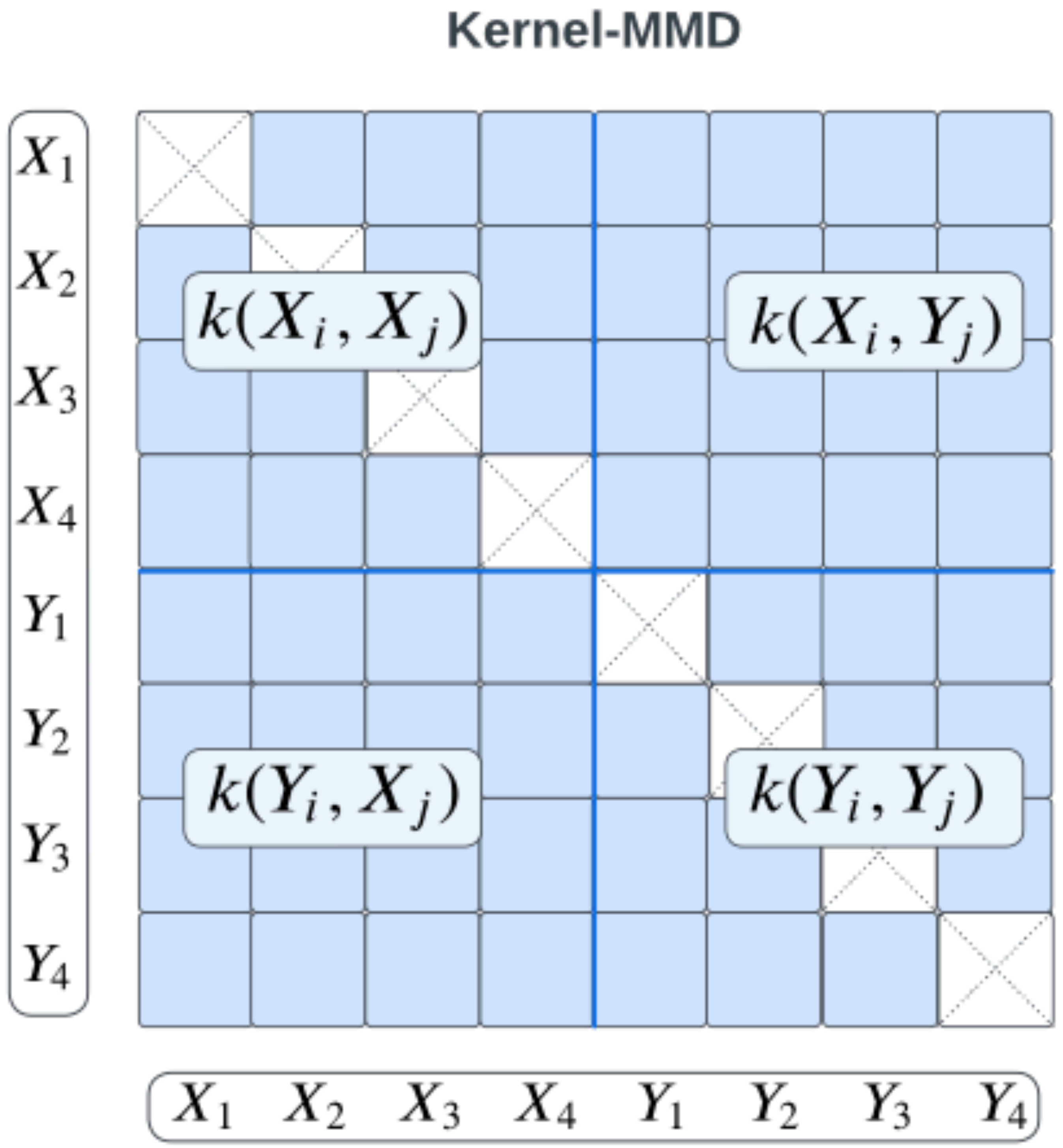
, and use an unbiased empirical estimator of MMD as

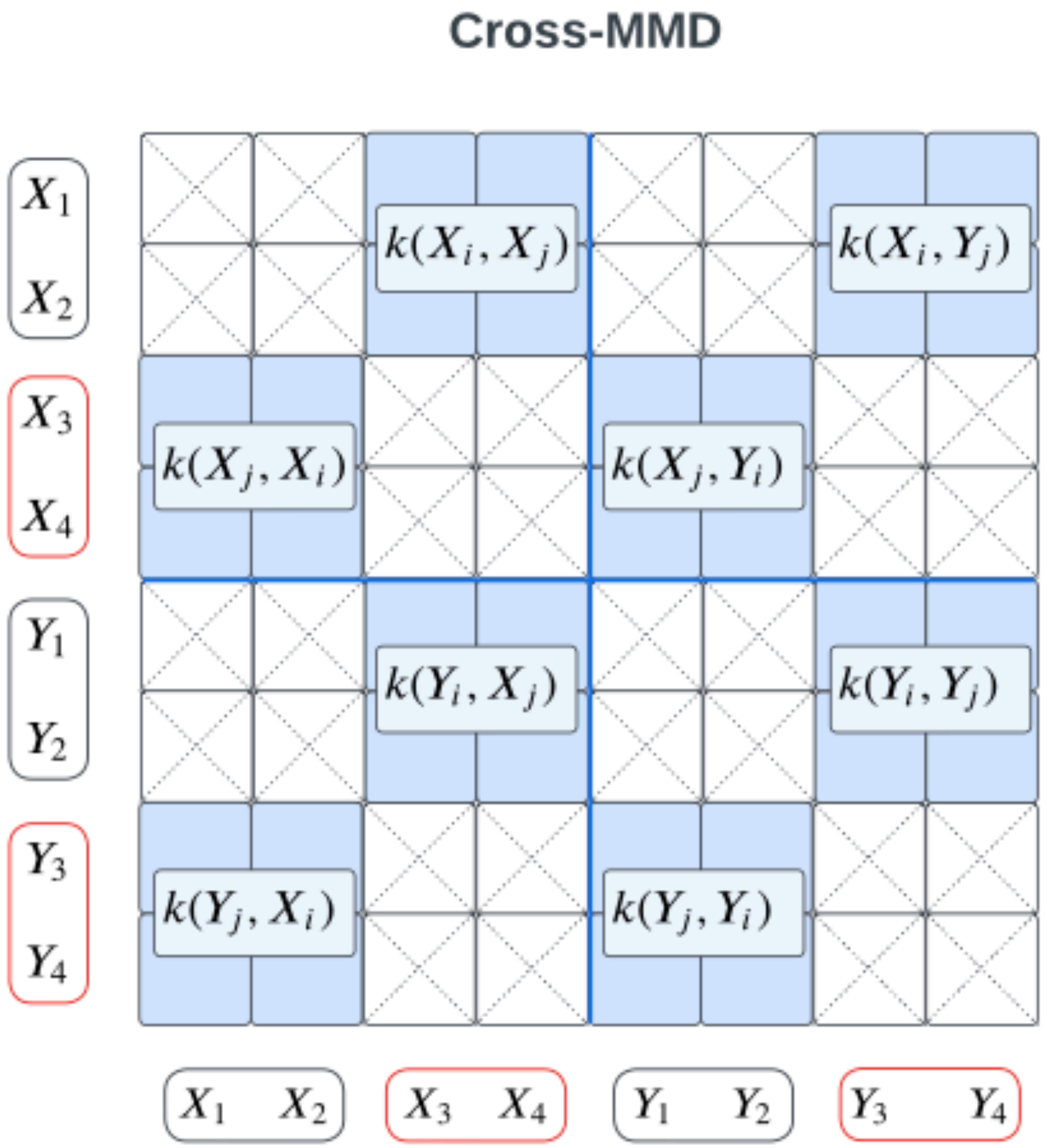
$$\mathrm{MMD}_u^2[\mathcal{F}, X, Y] = \frac{1}{m(m-1)} \sum_{i=1}^{m} \sum_{j \neq i}^{m} k(x_i, x_j) + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} k(y_i, y_j) - \frac{2}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} k(x_i, y_j)$$

- http://www.gatsby.ucl.ac.uk/~gretton/papers/columbia23/columbia23_21
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, *13*(1), 723-773.
- Murphy, K. P. (2018). Machine learning: A probabilistic perspective (adaptive computation and machine learning series). *The MIT Press: London, UK*.

Introduction

Instead of using permutation test, we use sample-splitting and studentization to handle this issue.



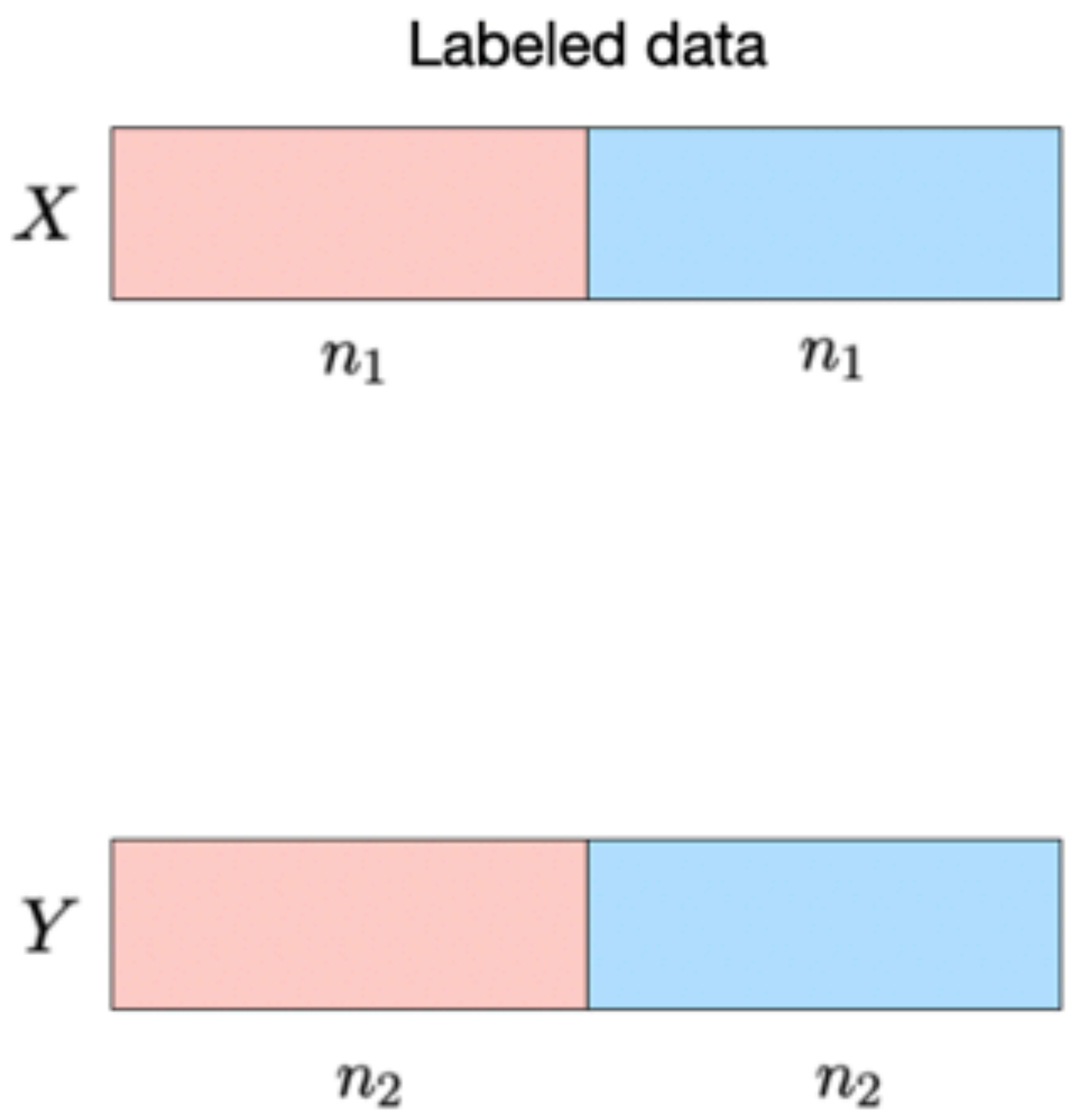


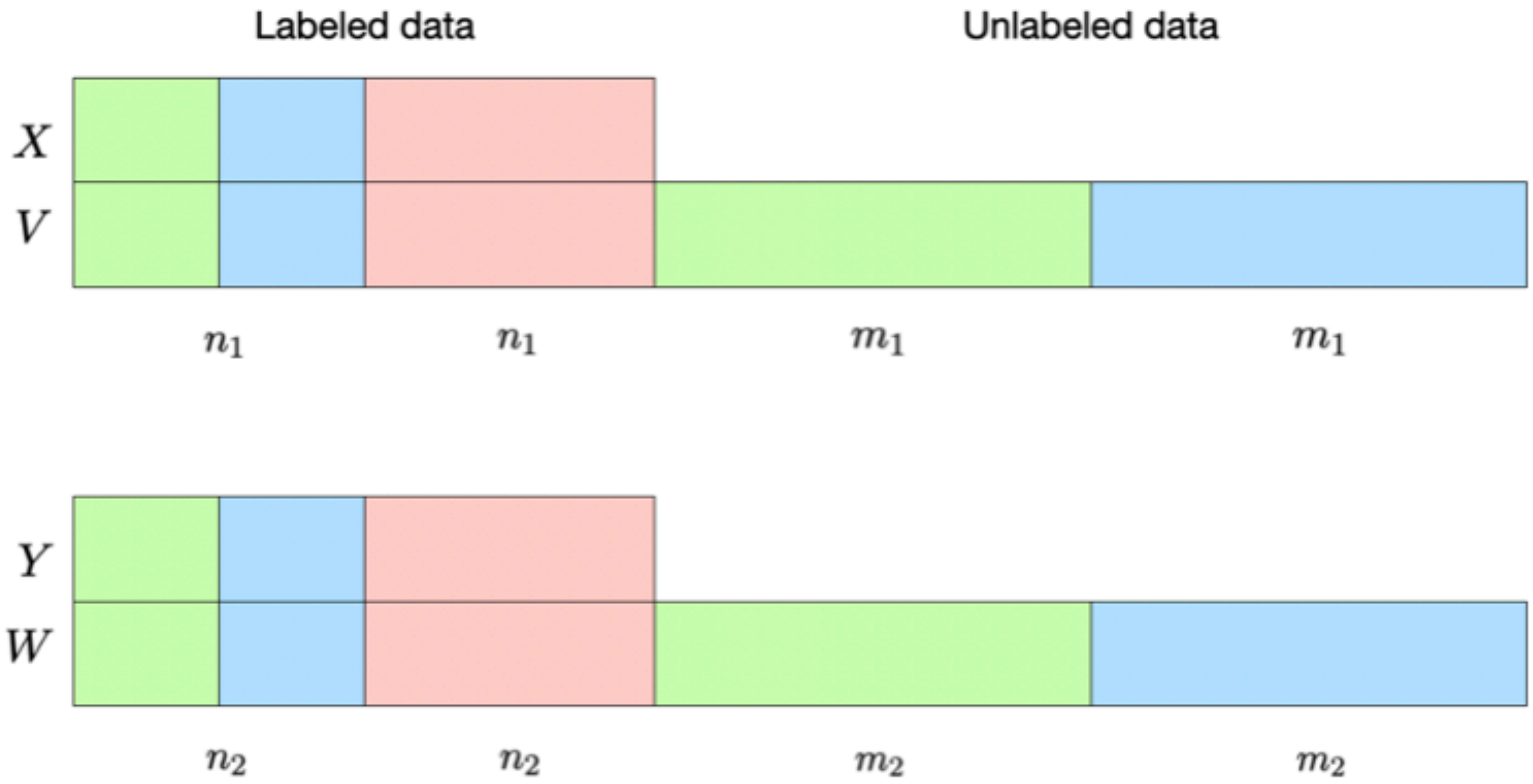
- Kim, I., & Ramdas, A. (2020). Dimension-agnostic inference using cross U-statistics. *arXiv preprint arXiv:2011.05068*.
- Shekhar, S., Kim, I., & Ramdas, A. (2022). A permutation-free kernel two-sample test. *Advances in Neural Information Processing Systems*, *35*, 18168-18180.

Method

- Setting

Now we have samples of both labeled and unlabeled data from two distributions





Method

- General Semi-Supervised Two-Sample Test

Estimate the witness function $\widehat{f}(\cdot)$ from $\mathcal{D}_{XV,2}$ and $\mathcal{D}_{YW,2}$

Then, from $\widehat{f}(X_1), \ldots, \widehat{f}(X_{n_1})$ and $\widehat{f}(Y_1), \ldots, \widehat{f}(Y_{n_2})$, we compute the estimator

$$\widehat{\mu}_{X,\widehat{f}} = \frac{1}{n_1}\sum_{i=1}^{n_1}\{\widehat{f}(X_i) - \mathbb{E}[\widehat{f}(X_i) \,|\, V_i, \widehat{f}]\} + \frac{1}{n_1+m_1}\sum_{i=1}^{n_1+m_1} \mathbb{E}[\widehat{f}(X_i) \,|\, V_i, \widehat{f}]$$

$$\widehat{\mu}_{Y,\widehat{f}} = \frac{1}{n_2}\sum_{i=1}^{n_2}\{\widehat{f}(Y_i) - \mathbb{E}[\widehat{f}(Y_i) \,|\, W_i, \widehat{f}]\} + \frac{1}{n_2+m_2}\sum_{i=1}^{n_2+m_2} \mathbb{E}[\widehat{f}(Y_i) \,|\, W_i, \widehat{f}].$$

and studentize it to obtain the test statistic $T^\star = \dfrac{\widehat{\mu}_{X,\widehat{f}} - \widehat{\mu}_{Y,\widehat{f}}}{\sqrt{\widehat{\sigma}^2_{X,\widehat{f}} + \widehat{\sigma}^2_{Y,\widehat{f}}}}$ where we reject the null if $T^\star > z_{1-\alpha} := \Phi^{-1}(1-\alpha)$

Method

- General Semi-Supervised Two-Sample Test

Estimate the witness function $\widehat{f}(\cdot)$ from $\mathcal{D}_{XV,2}$ and $\mathcal{D}_{YW,2}$

Then, from $\widehat{f}(X_1), \ldots, \widehat{f}(X_{n_1})$ and $\widehat{f}(Y_1), \ldots, \widehat{f}(Y_{n_2})$, we compute the estimator

$$\widehat{\mu}_{X,\widehat{f}} = \frac{1}{n_1}\sum_{i=1}^{n_1}\{\widehat{f}(X_i) - \mathbb{E}[\widehat{f}(X_i) \mid V_i, \widehat{f}]\} + \frac{1}{n_1+m_1}\sum_{i=1}^{n_1+m_1}\mathbb{E}[\widehat{f}(X_i) \mid V_i, \widehat{f}]$$

$$\widehat{\mu}_{Y,\widehat{f}} = \frac{1}{n_2}\sum_{i=1}^{n_2}\{\widehat{f}(Y_i) - \mathbb{E}[\widehat{f}(Y_i) \mid W_i, \widehat{f}]\} + \frac{1}{n_2+m_2}\sum_{i=1}^{n_2+m_2}\mathbb{E}[\widehat{f}(Y_i) \mid W_i, \widehat{f}].$$

and studentize it to obtain the test statistic $T^\star = \frac{\widehat{\mu}_{X,\widehat{f}} - \widehat{\mu}_{Y,\widehat{f}}}{\sqrt{\widehat{\sigma}^2_{X,\widehat{f}} + \widehat{\sigma}^2_{Y,\widehat{f}}}}$ where we reject the null if $T^\star > z_{1-\alpha} := \Phi^{-1}(1-\alpha)$

Method

- General Semi-Supervised Two-Sample Test

Estimate the witness function $\widehat{f}(\cdot)$ from $\mathcal{D}_{XV,2}$ and $\mathcal{D}_{YW,2}$

Then, from $\widehat{f}(X_1), \ldots, \widehat{f}(X_{n_1})$ and $\widehat{f}(Y_1), \ldots, \widehat{f}(Y_{n_2})$, we compute the estimator

$$\widehat{\mu}_{X,\widehat{f}} = \frac{1}{n_1}\sum_{i=1}^{n_1}\{\widehat{f}(X_i) - \mathbb{E}[\widehat{f}(X_i) \,|\, V_i, \widehat{f}]\} + \frac{1}{n_1+m_1}\sum_{i=1}^{n_1+m_1}\mathbb{E}[\widehat{f}(X_i) \,|\, V_i, \widehat{f}]$$

$$\widehat{\mu}_{Y,\widehat{f}} = \frac{1}{n_2}\sum_{i=1}^{n_2}\{\widehat{f}(Y_i) - \mathbb{E}[\widehat{f}(Y_i) \,|\, W_i, \widehat{f}]\} + \frac{1}{n_2+m_2}\sum_{i=1}^{n_2+m_2}\mathbb{E}[\widehat{f}(Y_i) \,|\, W_i, \widehat{f}].$$

and studentize it to obtain the test statistic $T^\star = \dfrac{\widehat{\mu}_{X,\widehat{f}} - \widehat{\mu}_{Y,\widehat{f}}}{\sqrt{\widehat{\sigma}^2_{X,\widehat{f}} + \widehat{\sigma}^2_{Y,\widehat{f}}}}$ where we reject the null if $T^\star > z_{1-\alpha} := \Phi^{-1}(1-\alpha)$

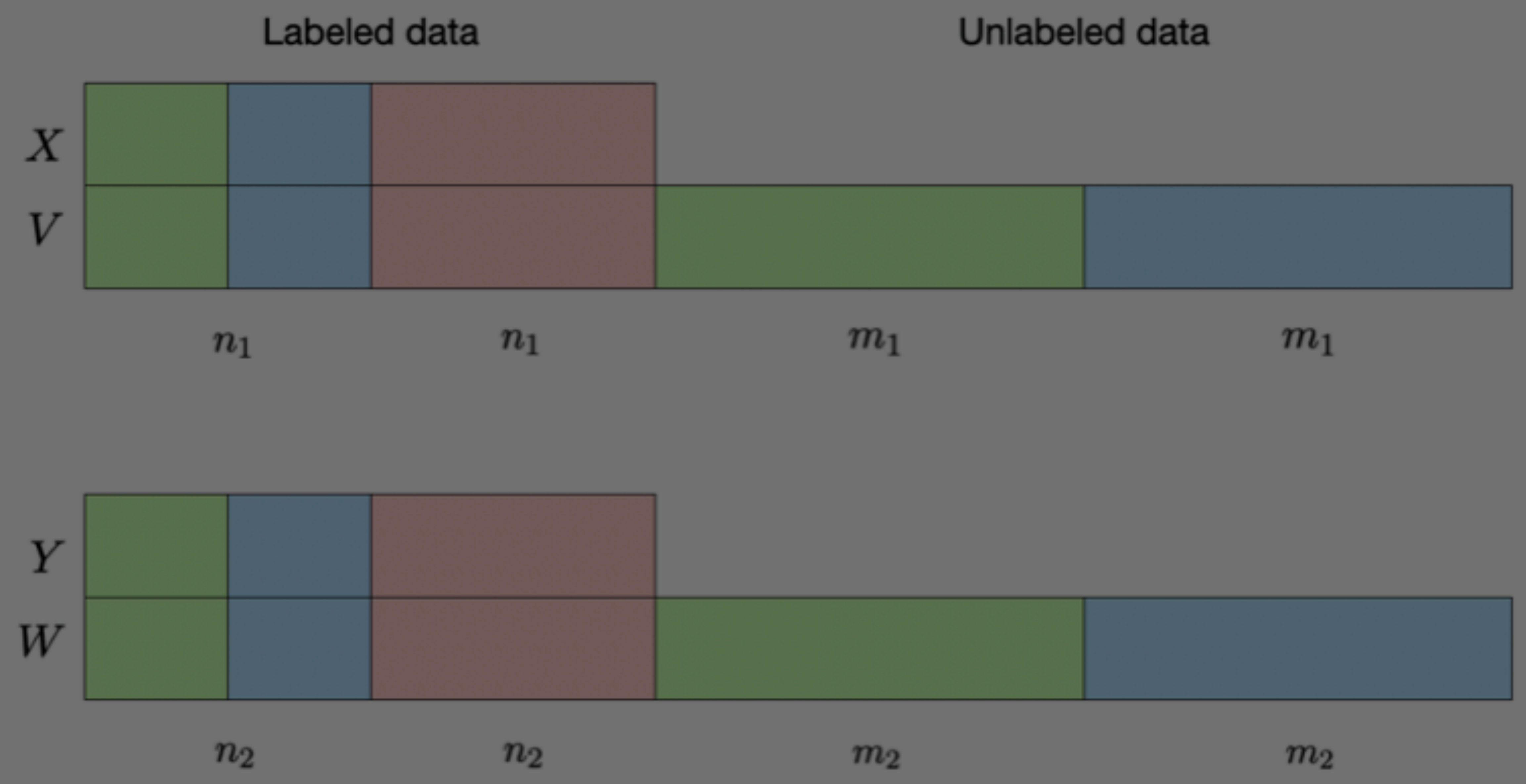
Method

- Semi-Supervised Kernel Two-Sample Test

Estimate the witness function

$$\widehat{f}(\cdot) = \frac{1}{n_1} \sum_{i \in \mathcal{I}_{XV,2}} k(X_i, \cdot) - \frac{1}{n_2} \sum_{i \in \mathcal{I}_{YW,2}} k(Y_i, \cdot)$$

Furthermore, estimate the conditional expectation $\mathbb{E}[\widehat{f}(X_i) \,|\, V_i, \widehat{f}]$ and $\mathbb{E}[\widehat{f}(Y_i) \,|\, W_i, \widehat{f}]$ using cross-fitting



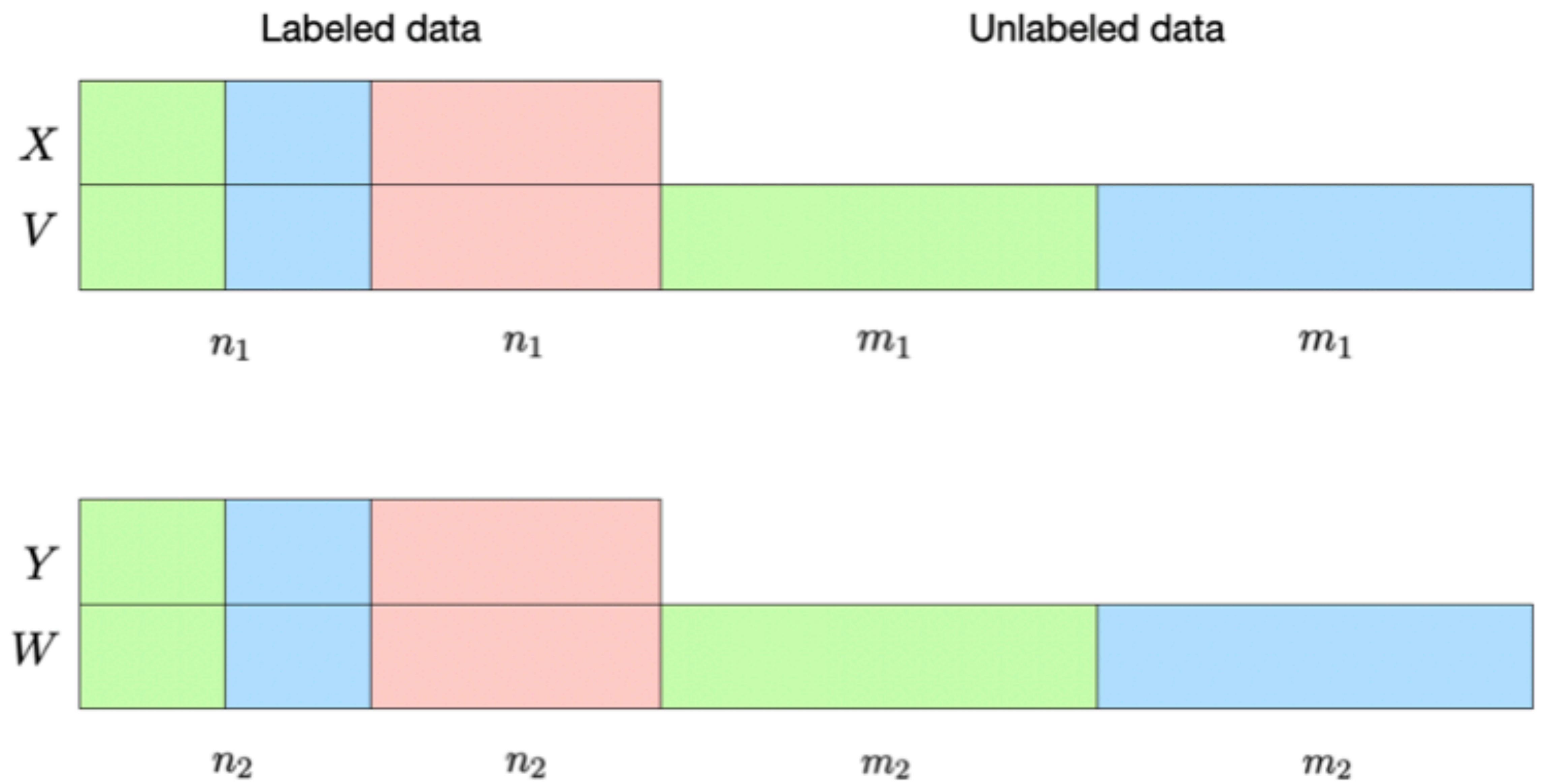
Method

- Semi-Supervised Kernel Two-Sample Test

Estimate the witness function

$$\widehat{f}(\cdot) = \frac{1}{n_1} \sum_{i \in \mathcal{I}_{XV,2}} k(X_i, \cdot) - \frac{1}{n_2} \sum_{i \in \mathcal{I}_{YW,2}} k(Y_i, \cdot)$$

Furthermore, estimate the conditional expectation $\mathbb{E}[\widehat{f}(X_i) \mid V_i, \widehat{f}]$ and $\mathbb{E}[\widehat{f}(Y_i) \mid W_i, \widehat{f}]$ using cross-fitting



Theoretical Results

1) Asymptotic normality under the null

- General Semi-Supervised Two-Sample Test

Theorem 1. *Suppose $X, Y \sim P$ where P is fixed and the kernel k may change as n increases. If $n_1 \asymp m_1$, $n_2 \asymp m_2$, and there exists $\delta > 0$ such that*

$$\mathbb{E}\big[\{\widehat{f}(X) - \mathbb{E}[\widehat{f}(X) \mid \widehat{f}]\}^{2+\delta}\big] / \sigma_{X,\widehat{f}}^{2+\delta} = o_P(n_1^{-\delta/2}),$$

$$\mathbb{E}\big[\{\widehat{f}(Y) - \mathbb{E}[\widehat{f}(Y) \mid \widehat{f}]\}^{2+\delta}\big] / \sigma_{Y,\widehat{f}}^{2+\delta} = o_P(n_2^{-\delta/2}),$$

then $T^\star \xrightarrow{d} N(0,1)$.

Theoretical Results

1) Asymptotic normality under the null

- Semi-Supervised Kernel Two-Sample Test

Theorem 2. *Suppose P is fixed and the kernel k changes as n increases. If $n_1 \asymp m_1$, $n_2 \asymp m_2$ and*

$$\lim_{n \to \infty} \frac{\mathbb{E}[\bar{k}^4(X_1, X_2)] \frac{n_1+n_2}{n_1 n_2} + \mathbb{E}[\bar{k}^2(X_1, X_3)\bar{k}^2(X_2, X_3)]}{\{\mathbb{E}[\bar{k}^2(X_1, X_2)]\}^2 \left(\frac{n_1 n_2}{n_1+n_2}\right)} = 0, \tag{3}$$

then $T^\star \xrightarrow{d} N(0,1)$.

where $\bar{k}(x,y) = k(x,y) - \mathbb{E}[k(X,Y) \,|\, X = x] - \mathbb{E}[k(X,Y) \,|\, Y = y] + \mathbb{E}[k(X,Y)]$

Theoretical Results

1) Asymptotic normality under the null

- Test statistic using cross-fitting

Theorem 3. *Suppose P and the kernel k are fixed, and $n_1 \asymp m_1$ and $n_2 \asymp m_2$. If the same moment condition (3) holds, and conditional expectations are estimated well, satisfying*

$$\frac{\mathbb{E}\big[\{\mathbb{E}[\widehat{f}(X)\,|\,V,\widehat{f}] - \widehat{\mathbb{E}}[\widehat{f}(X)\,|\,V,\widehat{f}]\}^2\,|\,\widehat{f}\big]}{\mathrm{Var}\{\widehat{f}(X)\,|\,\widehat{f}\}} = o_P(1),$$

$$\frac{\mathbb{E}\big[\{\mathbb{E}[\widehat{f}(Y)\,|\,W,\widehat{f}] - \widehat{\mathbb{E}}[\widehat{f}(Y)\,|\,W,\widehat{f}]\}^2\,|\,\widehat{f}\big]}{\mathrm{Var}\{\widehat{f}(Y)\,|\,\widehat{f}\}} = o_P(1),$$

then $\widehat{T} \xrightarrow{d} N(0,1)$.

Assuming linear estimator,

$$\sup_{i\geq 1} \mathbb{E}[\Delta_i^2] = o(1) \ \ \textit{for } \Delta_i = \mathbb{E}[\phi_i(X)\,|\,V] - \widehat{\mathbb{E}}[\phi_i(X)\,|\,V], \textit{ then } \widehat{T} \xrightarrow{d} N(0,1).$$

Theoretical Results

2) Asymptotic normality under the alternative

- Permutation-free & Semi-Supervised Kernel Two-Sample Test

Theorem 4. *Let P and Q be fixed distributions such that $P \neq Q$ with density functions p and q, respectively. Suppose that $\max\{\|p/q\|_\infty, \|q/p\|_\infty\} < C$ for some constant $C > 0$ and $n_1 \leq n_2$. If the centered kernel $\bar{k}_X$ and induced terms of $\bar{g}_X$ satisfies the moment conditoins of*

$$\frac{\mathbb{E}[\bar{k}_X(X_1, X_2)^4] + n_1 \mathbb{E}[\bar{k}_X(X_1, X_2)^2 \bar{k}_X(X_1, X_3)^2]}{n_1^2 \{\mathbb{E}[\bar{g}_X(X, X)]\}^2} = o(1), \quad \textit{and}$$

$$\frac{\mathrm{MMD}^4 \mathbb{E}[\bar{k}_X(X, X)^2]}{\{n_1 \mathbb{E}[\bar{g}_X(X, X)] + n_1^2 \mathbb{E}[\bar{g}_X(Y_1, Y_2)]\}^2} = o(1),$$

then $T_{c-mmd} \xrightarrow{d} N(0, 1)$.

Semi-Supervised Kernel Two-Sample Test

3) Asymptotic power expression with linear kernel

Theorem 5. *Suppose that assumption is fulfilled under the alternative. Then it holds that*

$$\mathbb{E}[\phi^\star] = \Phi\Big(z_\alpha + \frac{(\mu_X - \mu_Y)^\top(\mu_X - \mu_Y)}{\sqrt{\mathrm{tr}((\Lambda + \tilde{\Lambda})\Sigma)}}\Big) + o_P(1)$$

where $\Lambda = \frac{1}{n_1}\Sigma_{11} - \frac{m_1}{n_1(n_1+m_1)}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$, $\tilde{\Lambda} = \frac{1}{n_2}\tilde{\Sigma}_{11} - \frac{m_2}{n_2(n_2+m_2)}\tilde{\Sigma}_{12}\tilde{\Sigma}_{22}^{-1}\tilde{\Sigma}_{21}$, *and* $\Sigma = n_1^{-1}\Sigma_{11} + n_2^{-1}\tilde{\Sigma}_{11}$.

Semi-Supervised Kernel Two-Sample Test

3) Asymptotic power expression with linear kernel

- Using unlabeled data, semi-supervised test has higher power than the original kernel two-sample test in certain conditions.

$$\mathbb{E}[\phi_{perm}] = \Phi\big(z_\alpha + \frac{n(\mu_X - \mu_Y)^\top(\mu_X - \mu_Y)}{\sqrt{2\mathrm{tr}(\Sigma_{11}^2)}}\big) + o_P(1)$$

$$\mathbb{E}[\phi^\star] = \Phi\big(z_\alpha + \frac{n(\mu_X - \mu_Y)^\top(\mu_X - \mu_Y)}{\sqrt{4\mathrm{tr}(\Sigma_{11}^2) - 2\mathrm{tr}(\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11})}}\big) + o_P(1)$$

$$\mathbb{E}[\phi_{c-mmd}] = \Phi\big(z_\alpha + \frac{n(\mu_X - \mu_Y)^\top(\mu_X - \mu_Y)}{\sqrt{4\mathrm{tr}(\Sigma_{11}^2))}}\big) + o_P(1)$$

Numerical Analysis

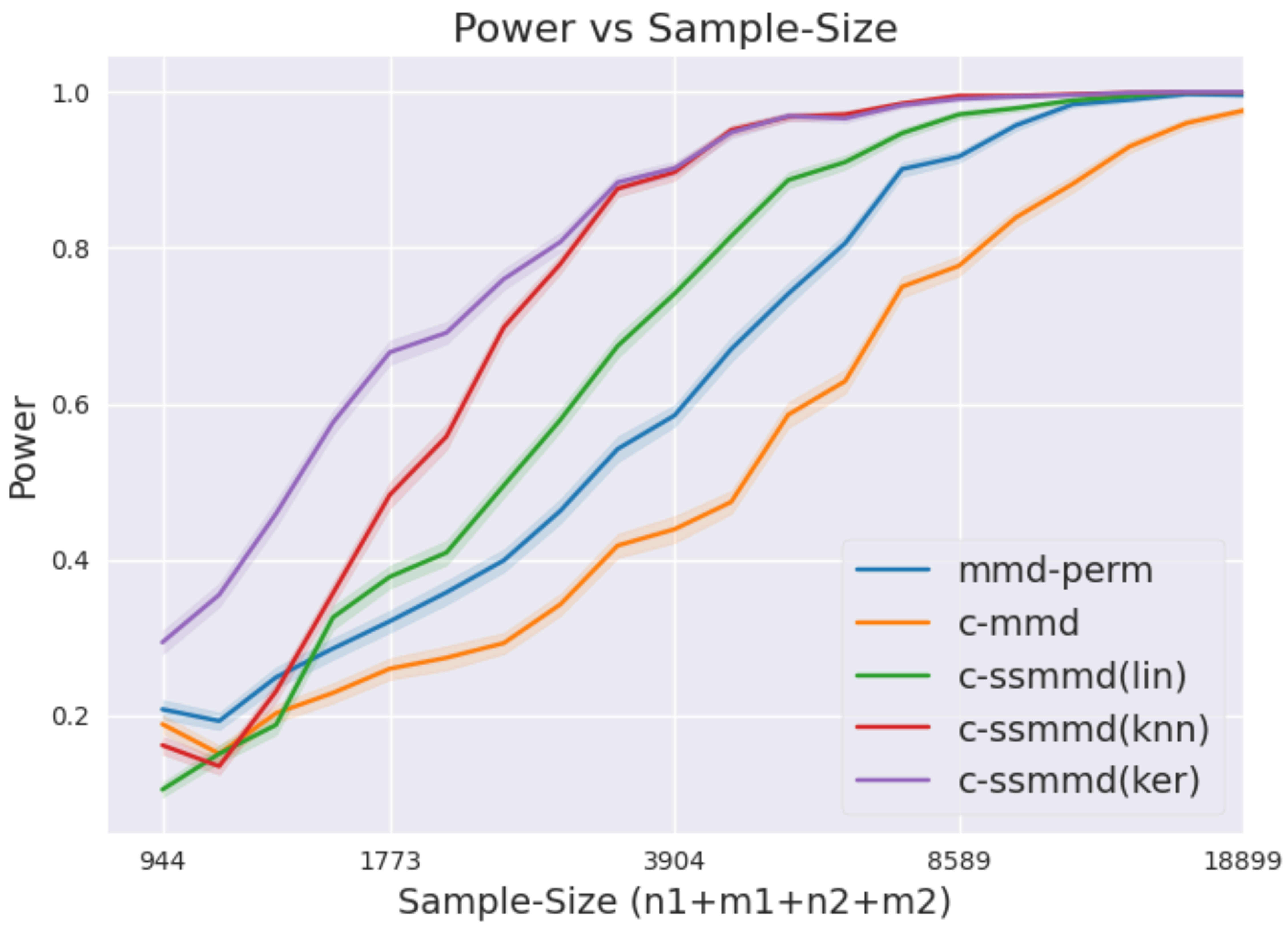
- Asymptotic Normality

$$X = V_{(1)} + V_{(2)} + 0.3Z$$
$$Y = W_{(1)} + W_{(2)} + 0.3Z$$
$$V \sim N_d(\mu_1, \Sigma_1), \quad \text{where} \quad \mu_1 = \mathbb{0}_d,\ \Sigma_1 = 0.3I_d + 0.7\mathbb{1}_d'\mathbb{1}_d$$
$$W \sim N_d(\mu_2, \Sigma_2), \quad \text{where} \quad \mu_2 = (\epsilon, ..., \epsilon)',\ \Sigma_2 = 0.3I_d + 0.7\mathbb{1}_d'\mathbb{1}_d$$

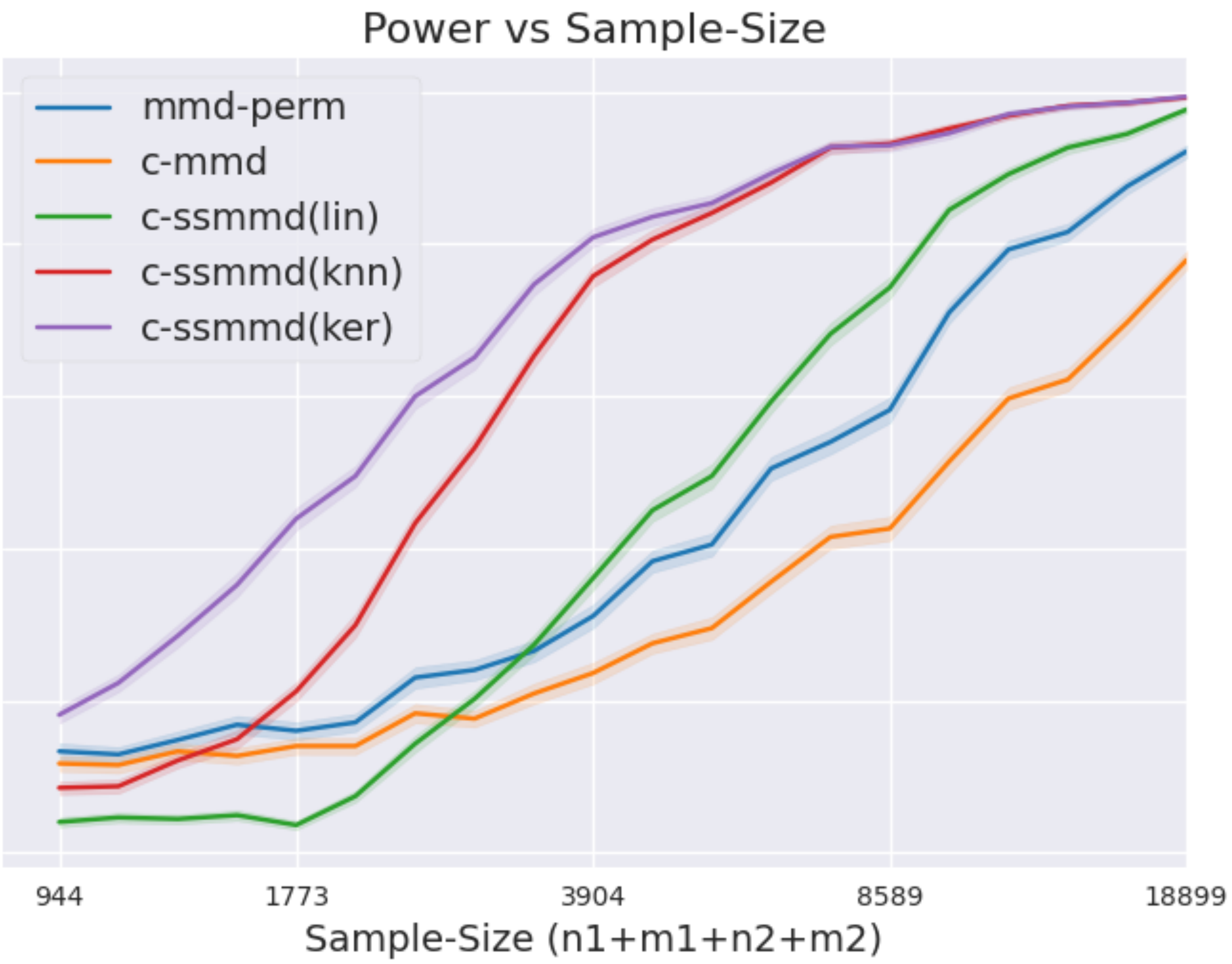
Numerical Analysis

- Comparison of power

$$X = V_{(1)} + \cdots + V_{(d)} + cZ$$
$$Y = W_{(1)} + \cdots + W_{(d)} + cZ$$
$$V \sim N_d(\mu_1, \Sigma_1), \quad \text{where} \quad \mu_1 = \mathbf{0}_d,\ \Sigma_1 = \rho \mathbb{1}_d \mathbb{1}_d^\top + (1-\rho) I_d$$
$$W \sim N_d(\mu_2, \Sigma_2), \quad \text{where} \quad \mu_2 = (\underbrace{\epsilon, ..., \epsilon}_{\text{p } elements}, 0, ..., 0)',\ \Sigma_2 = \rho \mathbb{1}_d \mathbb{1}_d^\top + (1-\rho) I_d$$



d=4,rho=0.9,pert=3,eps=0.5,n1=400,m1=8000,n2=500,m2=10000



d=10,rho=0.9,pert=5,eps=0.5,n1=400,m1=8000,n2=500,m2=10000

Conclusion

- Proposed semi-supervised two-sample testing, in particular, using kernel methods

- Proved asymptotic normality of the test statistic under the null and the alternative

- Derived asymptotic power expression of the power

Using unlabeled data is helpful in certain conditions!

Future Work

- Extension of witness function using unlabeled data when constructing itself

- Numerical Analysis for different cases

Reference

- Angelopoulos, A. N., Bates, S., Fannjiang, C., Jordan, M. I., & Zrnic, T. (2023). Prediction-powered inference. *Science*, *382*(6671), 669-674.
- Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, *13*(1), 723-773.
- Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
- Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R., & Jin, C. (2023). Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Computers in Biology and Medicine*, 107840.
- Liu, F., Xu, W., Lu, J., Zhang, G., Gretton, A., & Sutherland, D. J. (2020, November). Learning deep kernels for non-parametric two-sample tests. In *International conference on machine learning* (pp. 6316-6326). PMLR.
- Murphy, K. P. (2018). Machine learning: A probabilistic perspective (adaptive computation and machine learning series). *The MIT Press: London, UK*.
- Kim, I., & Ramdas, A. (2020). Dimension-agnostic inference using cross U-statistics. arXiv preprint arXiv:2011.05068.
- Shekhar, S., Kim, I., & Ramdas, A. (2022). A permutation-free kernel two-sample test. *Advances in Neural Information Processing Systems*, *35*, 18168-18180.
- Tony Cai, T., & Guo, Z. (2020). Semisupervised inference for explained variance in high dimensional linear regression and its applications. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *82*(2), 391-419.
- Yongho, Jeon (2023), Lecture notes for nonparametric function estimation
- Zhang, A., Brown, L. D., & Cai, T. T. (2019). Semi-supervised inference: General theory and estimation of means.
- Zhang, Y., & Bradic, J. (2022). High-dimensional semi-supervised learning: in search of optimal inference of the mean. *Biometrika*, *109*(2), 387-403.
- Zhu, B., Ding, M., Jacobson, P., Wu, M., Zhan, W., Jordan, M., & Jiao, J. (2024). Doubly-Robust Self-Training. *Advances in Neural Information Processing Systems*, *36*.
- https://towardsdatascience.com/supervised-vs-unsupervised-learning-in-2-minutes-72dad148f242
- http://www.gatsby.ucl.ac.uk/~gretton/papers/columbia23/columbia23_21