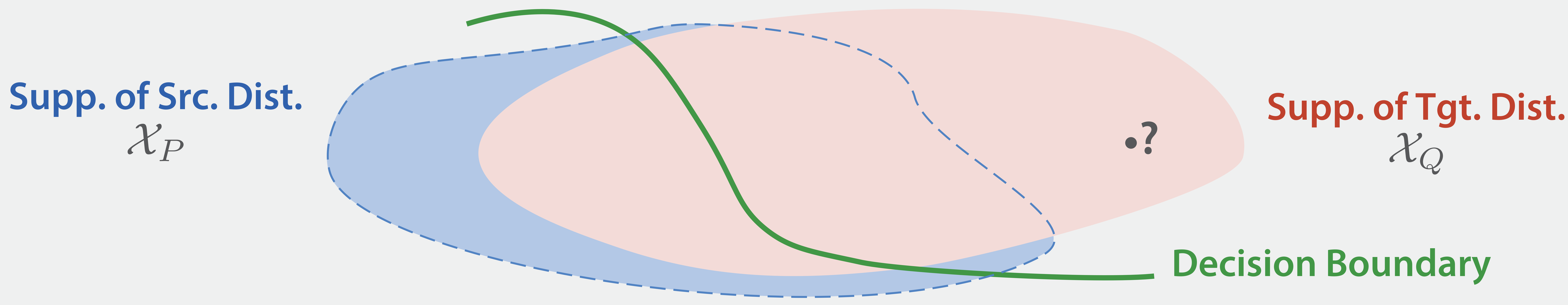
Harnessing the Power of Vicinity-Informed Analysis for Classification under Covariate Shift

Mitsuhiro Fujikawa[1,3], Youhei Akimoto[1,3], Jun Sakuma[2,3], and Kazuto Fukuchi[1,3]

[1]University of Tsukuba [2]Institute of Science Tokyo [3]RIKEN AIP

arXiv

Establishing transfer learning “success” under support non-containment environments



Binary Classification under Covariate Shift

Input $X \in \mathcal{X}$ with metric space $(\mathcal{X}, \rho)$ and label $Y \in \{0, 1\}$

$P^{n_P} \sim$ Src. sample $\{(X_i, Y_i)\}_{i=1}^{n_P}$ → Learner

$Q^{n_Q} \sim$ Tgt. sample $\{(X_i, Y_i)\}_{i=n_P+1}^{n_Q}$ → Learner

Learner → Classifier $\hat{h}$

Test data $(X, Y) \sim Q$

Error $err_Q(\hat{h}) = \mathbb{E}_Q \mathbb{1}\{\hat{h}(X) \neq Y\}$

Covariate shift

$$P_{Y|X}(Y=1|X) = Q_{Y|X}(Y=1|X) = \eta(X) \quad \& \quad P_X \neq Q_X$$

Basic assumptions

Smoothness $|\eta(x) - \eta(x')| \leq C_\alpha \rho^\alpha(x, x')$ Noise condition $Q_X(0 < |\eta(X) - \frac{1}{2}| < t) \leq C_\beta t^\beta$

When is transfer learning deemed successful?

Success = **Source sample-size consistency**

$$\sup_{P,Q \in \mathcal{P}} \mathbf{E}[\mathcal{E}_Q(\hat{h})] \to 0 \text{ as } n_Q \to \infty$$

Excess Error $\mathcal{E}_Q(h) = err_Q(h) - \inf_{h^*:\text{measuable}} err_Q(h^*)$

Support non-containment environments

$$\mathcal{X}_Q \not\subseteq \mathcal{X}_P$$

Dissimilarity measures

Kpotufe et al. (2021) [reinterpretation]

$$\Delta_{\mathrm{KM}}(P, Q; r) = \sup_{x \in \mathcal{X}_Q} \frac{Q_X(B(x,r))}{P_X(B(x,r))} \qquad \Delta_{\mathrm{DM}}(Q, Q; r) = \sup_{x \in \mathcal{X}_Q} \frac{1}{Q_X(B(x,r))}$$

$$\Delta_{\mathrm{BCN}}(Q, Q; r) = \mathcal{N}(\mathcal{X}_Q, \rho, r)$$

Pathak et al. (2022)

$$\Delta_{\mathrm{PMW}}(P, Q; r) = \int \frac{1}{P_X(B(x,r))} Q_X(dx)$$

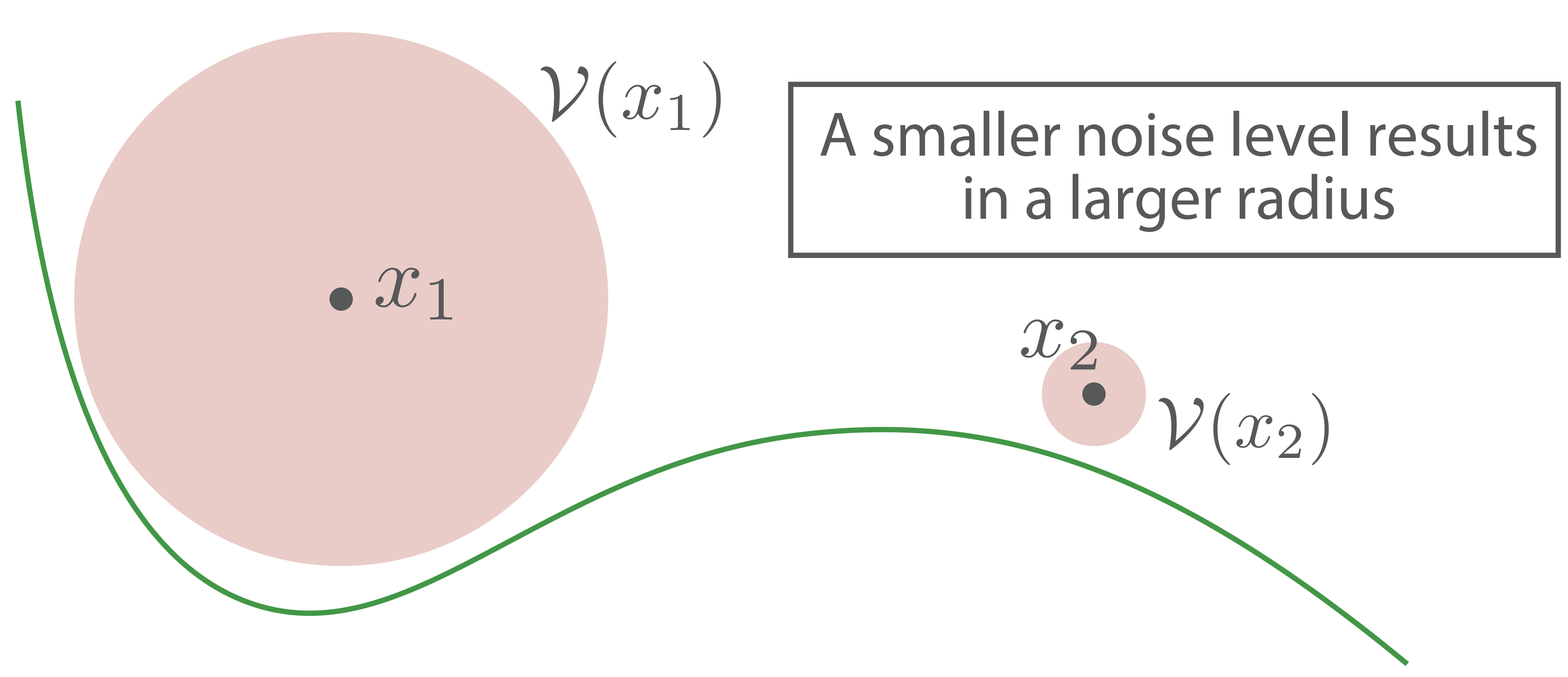
Become **infinity** in support non-containment environment

Our

$$\Delta_{\mathcal{V}}(P, Q; r) = \int \inf_{x' \in \mathcal{V}(x)} \frac{1}{P_X(B(x',r))} Q_X(dx)$$

Avoid to become infinity by taking infimum over vicinity set

$$\mathcal{V}(x) = \left\{ x' \in \mathcal{X} : 2C_\alpha \rho^\alpha(x, x') < \left| \eta(x) - \frac{1}{2} \right| \right\} \cup \{x\}$$



Transfer- and Self-exponents

transfer-exponent of (P, Q) is τ if $\Delta(P, Q; r) = O(r^{-\tau})$

self-exponent of Q is ψ if $\Delta(Q, Q; r) = O(r^{-\psi})$

	τ	ψ
Kpotufe et al. (2021)	$\tau_{\Delta_{\mathrm{KM}}} + \min\{\psi_{\Delta_{\mathrm{DM}}}, \psi_{\Delta_{\mathrm{BCN}}}\}$	$\min\{\psi_{\Delta_{\mathrm{DM}}}, \psi_{\Delta_{\mathrm{BCN}}}\}$
Pathak et al. (2022)	$\tau_{\Delta_{\mathrm{PMW}}}$	$\psi_{\Delta_{\mathrm{PMW}}}$
our	$\tau_{\Delta_{\mathcal{V}}}$	$\psi_{\Delta_{\mathcal{V}}}$

τ_Δ and ψ_Δ are minimum exponents (P, Q) has

Universal rate

k-NN classifier w/ an appropriate *k* achieves

$$\mathbf{E}[\mathcal{E}_Q(\hat{h})] \leq C \left(n_P^{\frac{1+\beta}{2+\beta+\max\{1,\tau/\alpha\}}} + n_P^{\frac{1+\beta}{2+\beta+\max\{1,\psi/\alpha\}}} \right)^{-1}$$

Non-transfer rate $n^{-\frac{1+\beta}{2+\beta+d/\alpha}}$

Comparision of rates

For **any** (P, Q)

$$\tau_{\Delta_{\mathcal{V}}} \leq \tau_{\Delta_{\mathrm{PMW}}} \leq \tau_{\Delta_{\mathrm{KM}}} + \min\{\psi_{\Delta_{\mathrm{DM}}}, \psi_{\Delta_{\mathrm{BCN}}}\}$$

$$\psi_{\Delta_{\mathcal{V}}} \leq \psi_{\Delta_{\mathrm{PMW}}} \leq \min\{\psi_{\Delta_{\mathrm{DM}}}, \psi_{\Delta_{\mathrm{BCN}}}\}$$

Instance-wise supriority

k-NN classifier

Src. sample $\{(X_i, Y_i)\}_{i=1}^{n_P}$ Tgt. sample $\{(X_i, Y_i)\}_{i=n_P+1}^{n_Q}$

↓

Combined sample $\{(X_i, Y_i)\}_{i=1}^{n_P+n_Q}$

Est. reg. func. $\hat{\eta}_k(X) = \frac{1}{k} \sum_{i=1}^{k} Y_{(i)}$

Experiments

$$p_X(x) \propto (1 - x^2)^{-\tau/2}$$

$$\mathcal{X}_P = \left[-\frac{8^{1/\alpha} \cdot 2 - 1}{8^{1/\alpha} \cdot 2}, \frac{8^{1/\alpha} \cdot 2 - 1}{8^{1/\alpha} \cdot 2}\right]$$

$$\mathcal{X}_Q = [-1, 1]$$

$$\eta(x) = \frac{1}{2} + \frac{1}{2}\mathrm{sgn}(x)|x|^\alpha$$