Towards Accurate Model Selection in Deep Unsupervised Domain Adaptation

Kaichao You[1], Ximei Wang[1], Mingsheng Long[1], Michael I. Jordan[2]

[1]School of Software, Tsinghua University
[1]National Engineering Lab for Big Data Software
[2]University of California, Berkeley

International Conference on Machine Learning ICML 2019

Outline

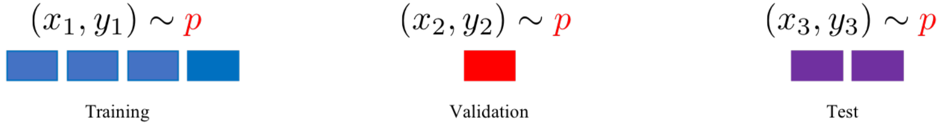
Validation in UDA: the problem

- Supervised Learning

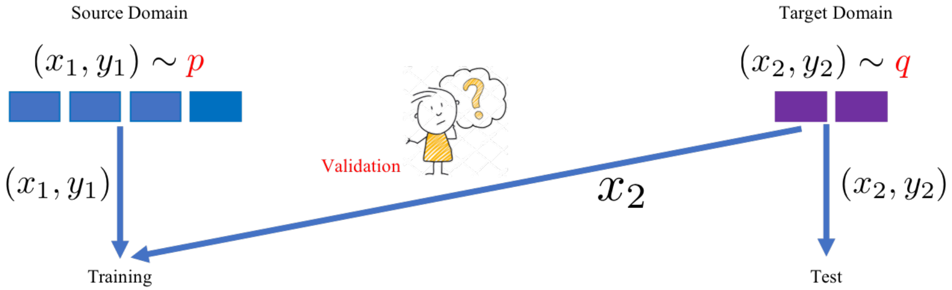
$(x_1, y_1) \sim p$ $\qquad$ $(x_2, y_2) \sim p$ $\qquad$ $(x_3, y_3) \sim p$

Validation in UDA: the problem

- Supervised Learning



- Unsupervised Domain Adaptation



Outline

IWCV: the previous solution

- Covariate Shift Assumption $p(y|\mathbf{x}) = q(y|\mathbf{x})$

IWCV: the previous solution

- Covariate Shift Assumption $p(y|\mathbf{x}) = q(y|\mathbf{x})$
- Model Selection: estimate Target Risk $\mathcal{R}(g) = \mathbb{E}_{\mathbf{x}\sim q}\ell(g(\mathbf{x}), y)$

IWCV: the previous solution

- Covariate Shift Assumption $p(y|\mathbf{x}) = q(y|\mathbf{x})$
- Model Selection: estimate Target Risk $\mathcal{R}(g) = \mathbb{E}_{\mathbf{x}\sim q}\ell(g(\mathbf{x}), y)$
- Importance Weighted Cross Validation [1]

$$\mathbb{E}_{\mathbf{x}\sim p}w(\mathbf{x})\ell(g(\mathbf{x}), y) = \mathbb{E}_{\mathbf{x}\sim p}\frac{q(\mathbf{x})}{p(\mathbf{x})}\ell(g(\mathbf{x}), y) = \mathbb{E}_{\mathbf{x}\sim q}\ell(g(\mathbf{x}), y) = \mathcal{R}(g)$$

[1] Covariate shift adaptation by importance weighted cross validation, JMLR'2007

IWCV: the previous solution

- Covariate Shift Assumption $p(y|\mathbf{x}) = q(y|\mathbf{x})$
- Model Selection: estimate Target Risk $\mathcal{R}(g) = \mathbb{E}_{\mathbf{x}\sim q}\ell(g(\mathbf{x}), y)$
- Importance Weighted Cross Validation [1]

$$\mathbb{E}_{\mathbf{x}\sim p} w(\mathbf{x})\ell(g(\mathbf{x}), y) = \mathbb{E}_{\mathbf{x}\sim p}\frac{q(\mathbf{x})}{p(\mathbf{x})}\ell(g(\mathbf{x}), y) = \mathbb{E}_{\mathbf{x}\sim q}\ell(g(\mathbf{x}), y) = \mathcal{R}(g)$$

  - Unbiased but the variance is unbounded
  - Density ratio is not readily accessible

[1] Covariate shift adaptation by importance weighted cross validation, JMLR'2007

Outline

Deep Embedded Validation

- IWCV's variance[1]: $\mathrm{Var}_{\mathbf{x}\sim p}[\ell_w] \le d_{\alpha+1}(q\|p)\, \mathcal{R}(g)^{1-\frac{1}{\alpha}} - \mathcal{R}(g)^2$.

[1] Learning Bounds for Importance Weighting, NeurIPS'2010

Deep Embedded Validation

- IWCV's variance[1]: $\mathrm{Var}_{\mathbf{x}\sim p}[\ell_w] \leq d_{\alpha+1}(q\|p)\ \mathcal{R}(g)^{1-\frac{1}{\alpha}} - \mathcal{R}(g)^2$.
- Feature adaptation reduces distribution discrepancy[2]

[1]Learning Bounds for Importance Weighting, NeurIPS'2010
[2]Conditional Adversarial Domain Adaptation, NeurIPS'2018

Deep Embedded Validation

- IWCV's variance[1]: $\mathrm{Var}_{\mathbf{x}\sim p}[\ell_w] \leq d_{\alpha+1}(q\|p)\ \mathcal{R}(g)^{1-\frac{1}{\alpha}} - \mathcal{R}(g)^2$.
- Feature adaptation reduces distribution discrepancy[2]
- Control variate explicitly reduces the variance
  - $\mathbb{E}[z] = \zeta, \mathbb{E}[t] = \tau$
  - $z^\star = z + \eta(t - \tau)$.
  - $\mathbb{E}[z^\star] = \mathbb{E}[z] + \eta\mathbb{E}[t - \tau] = \zeta + \eta(\mathbb{E}[t] - \mathbb{E}[\tau]) = \zeta$.
  - $\mathrm{Var}[z^\star] = \mathrm{Var}[z + \eta(t - \tau)] = \eta^2\mathrm{Var}[t] + 2\eta\mathrm{Cov}(z, t) + \mathrm{Var}[z]$
  - $\min \mathrm{Var}[z^\star] = (1 - \rho_{z,t}^2)\mathrm{Var}[z]$, when $\hat{\eta} = -\frac{\mathrm{Cov}(z,t)}{\mathrm{Var}[t]}$

[1]Learning Bounds for Importance Weighting, NeurIPS'2010
[2]Conditional Adversarial Domain Adaptation, NeurIPS'2018

Deep Embedded Validation

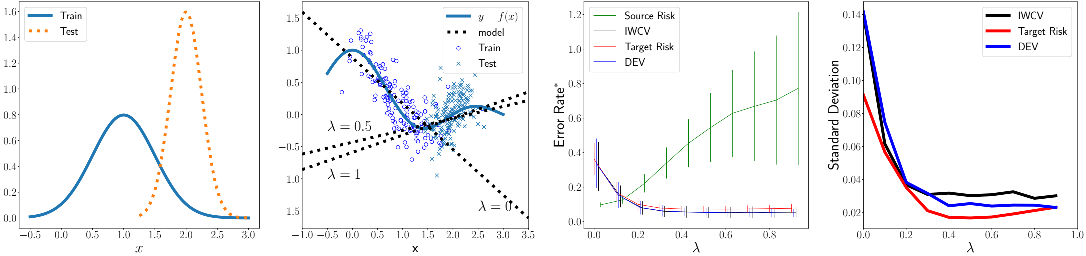
- IWCV's variance[1]: $\mathrm{Var}_{\mathbf{x}\sim p}[\ell_w] \leq d_{\alpha+1}(q\|p)\ \mathcal{R}(g)^{1-\frac{1}{\alpha}} - \mathcal{R}(g)^2$.
- Feature adaptation reduces distribution discrepancy[2]
- Control variate explicitly reduces the variance
  - $\mathbb{E}[z] = \zeta, \mathbb{E}[t] = \tau$
  - $z^\star = z + \eta(t - \tau)$.
  - $\mathbb{E}[z^\star] = \mathbb{E}[z] + \eta\mathbb{E}[t - \tau] = \zeta + \eta(\mathbb{E}[t] - \mathbb{E}[\tau]) = \zeta$.
  - $\mathrm{Var}[z^\star] = \mathrm{Var}[z + \eta(t - \tau)] = \eta^2\mathrm{Var}[t] + 2\eta\mathrm{Cov}(z, t) + \mathrm{Var}[z]$
  - $\min \mathrm{Var}[z^\star] = (1 - \rho_{z,t}^2)\mathrm{Var}[z]$, when $\hat{\eta} = -\frac{\mathrm{Cov}(z,t)}{\mathrm{Var}[t]}$
- Density ratio can be estimated discriminatively.[3]

[1] Learning Bounds for Importance Weighting, NeurIPS'2010
[2] Conditional Adversarial Domain Adaptation, NeurIPS'2018
[3] Discriminative learning for differing training and test distributions, ICML'2007

Outline

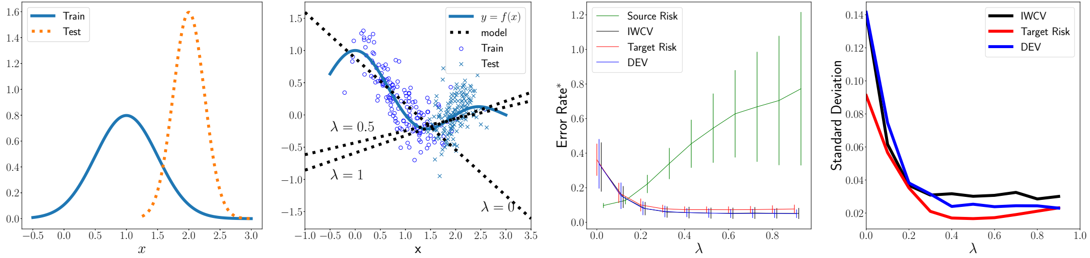
Experiments

- Experiments on a toy problem under covariate shift



Experiments

- Experiments on a toy problem under covariate shift



- Experiments on real-world problems
    - Various datasets: VisDA/Office/Digits
    - Various models: CDAN, MCD, GTA
    - Deep Embedded Validation is empirically validated ☺

Thanks!

Code available at github.com/thuml/Deep-Embedded-Validation
Poster: tonight at Pacific Ballroom #259