Towards Accurate Model Selection in Deep Unsupervised Domain Adaptation

Kaichao You\textsuperscript{1}, Ximei Wang\textsuperscript{1}, Mingsheng Long\textsuperscript{1}, Michael I. Jordan\textsuperscript{2}

\textsuperscript{1}School of Software, Tsinghua University
\textsuperscript{1}National Engineering Lab for Big Data Software
\textsuperscript{2}University of California, Berkeley

International Conference on Machine Learning ICML 2019
Outline

1 Validation in UDA: the problem

2 IWCV: the previous solution

3 Deep Embedded Validation

4 Experiments
Validation in UDA: the problem

- Supervised Learning
  \[(x_1, y_1) \sim p\]
  
  Training

- Validation
  \[(x_2, y_2) \sim p\]
  
  Validation

- Test
  \[(x_3, y_3) \sim p\]
  
  Test
Validation in UDA: the problem

- **Supervised Learning**
  \[(x_1, y_1) \sim p\]  
  \[(x_2, y_2) \sim p\]  
  \[(x_3, y_3) \sim p\]

  - Training
  - Validation
  - Test

- **Unsupervised Domain Adaptation**
  \[(x_1, y_1) \sim p\]

  - Source Domain
  - Validation
  - Target Domain
  - Test
Outline

1 Validation in UDA: the problem

2 IWCV: the previous solution

3 Deep Embedded Validation

4 Experiments
IWCV: the previous solution

- Covariate Shift Assumption $p(y|x) = q(y|x)$
IWCV: the previous solution

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

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

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

$^1$Covariate shift adaptation by importance weighted cross validation, JMLR’2007
IWCV: the previous solution

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

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

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

\(^1\)Covariate shift adaptation by importance weighted cross validation, JMLR’2007
Outline

1. Validation in UDA: the problem
2. IWCV: the previous solution
3. Deep Embedded Validation
4. Experiments
Deep Embedded Validation

- IWCV’s variance\(^1\): \( \text{Var}_{x \sim p}[\ell_w] \leq d^{\alpha+1}(q\|p) \, R(g)^{1-\frac{1}{\alpha}} - R(g)^2 \).

\(^1\)Learning Bounds for Importance Weighting, NeurIPS’2010
Deep Embedded Validation

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

---
\(^1\)Learning Bounds for Importance Weighting, NeurIPS’2010
\(^2\)Conditional Adversarial Domain Adaptation, NeurIPS’2018
Deep Embedded Validation

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

1 Validation in UDA: the problem
2 IWCV: the previous solution
3 Deep Embedded Validation
4 Experiments
Experiments

- Experiments on a toy problem under covariate shift

![Graphs showing error rate and standard deviation for different λ values.](image)

Kaichao You et al. Deep Embedded Validation June 12, 2019 9 / 10
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