On cross-validation under covariate shift

WM Kouw M Loog

Abstract

Standard cross-validation for L^2 regularization parameter estimation is suboptimal in a covariate shift setting, because it does not account for differences between the training (source domain) and test (target domain) data. Assigning importance weights to the source validation data scales the source validation risk to match the target risk and produces closer-to-optimal estimates. However, results of an experiment with a diverse set of importance weight estimators shows that importance weighted cross-validation consistently underestimates the optimal target regularization parameter.

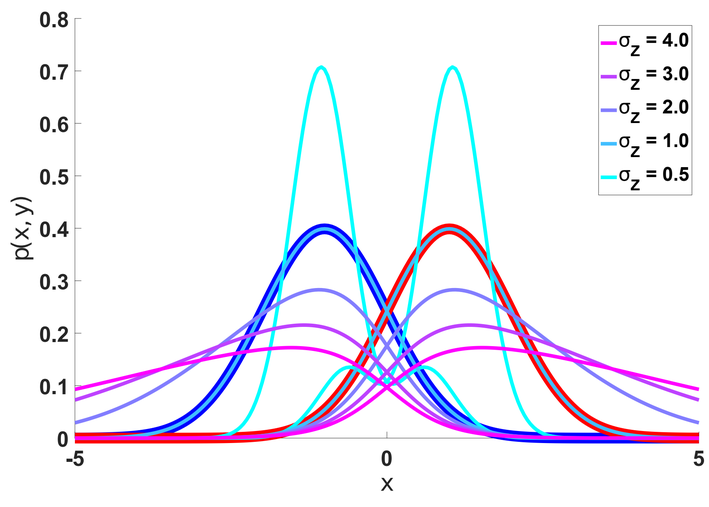
Covariate shift

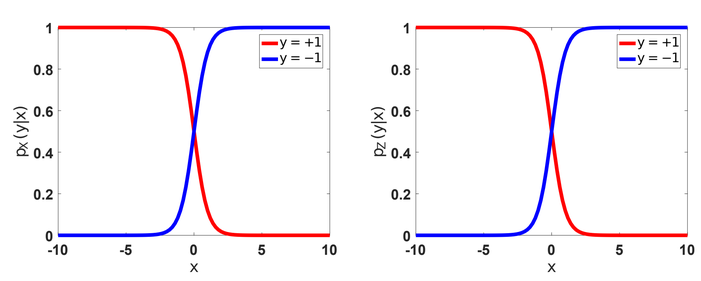
Domain adaptation is the supervised learning setting, where the marginal data distributions of the training data (source domain) and test data (target domain) differ:

$$p_{\mathcal{X}}(x) \neq p_{\mathcal{Z}}(x)$$

But the class-posterior distributions are the same:

$$p_{\mathcal{Y}|\mathcal{X}}(y \mid x) = p_{\mathcal{Y}|\mathcal{Z}}(y \mid x)$$





Problem

The mean squared error curve of a source least-squares classifier as a function of the regularization parameter shifts for different target domain variances. It shows that the optimal value of the L^2 regularization parameter depends on the domain dissimilarity.

$$\mathrm{MSE}(\lambda) = \frac{1}{m}\sum_{j=1}^{m}\left(Z_j\left[(X^\top X + \lambda I)^{-1}X^\top y\right] - u_j\right)^2$$

Importance weight estimators

Cross-validation can be partially corrected through importance weighing the source validation data. By matching the validation data to the target data, the shift in the MSE curve is reduced.

Importance weights are usually estimated through comparing the discrepancy between the data marginal distributions.

$$\hat{w}_{rG} = \frac{\mathcal{N}(x \mid \hat{\mu}_T, \hat{\sigma}_T^2)}{\mathcal{N}(x \mid \hat{\mu}_S, \hat{\sigma}_S^2)}$$

$$\hat{w}_{kliep} = \arg\max_{w \in W} \sum_{j=1}^{m} \log \sum_{i=1}^{n} w_i K(x_i, z_j)$$

$$s.t. \sum_{i=1}^{n} w_i K(x_i, z_j) = n$$

$$\hat{w}_{kmm} = \arg\min_{w \in W} \frac{1}{2} w^\top K(x, x') w - \frac{n}{m}\sum_{j=1}^{m} K(x, z_j)^\top w$$

$$s.t.\ w_i \in [0 \quad B]$$

$$\left|\frac{1}{n}\sum_{i=1}^{n} w_i - 1\right| \leq \epsilon$$

$$\hat{w}_{nn} = \left|C_i \cap \{z_j\}_{j=1}^{m}\right| + 1$$

Experiment

Evaluating the importance weight estimators as well as the true weights shows that importance weighted cross-validation still consistently underestimates the value of the optimal target regularization parameter.

$$\hat{\lambda} = \arg\min_{\lambda \in \Lambda} \frac{1}{|X_V|}\sum_{i=1}^{|X_V|} w_i\left(X_{Vi}\left[(X_T^\top X_T + \lambda I)^{-1}X_T^\top y_T\right] - y_{Vi}\right)^2$$

DISCUSSION

Our experiment shows that both the standard cross-validation and importance weighted cross-validation procedure underestimate the optimal value for the L^2 regularization parameter. Furthermore, the bias seems to be a function of the domain dissimilarity. Future work will aim to characterize the bias exactly.

TU Delft

prlab.tudelft.nl
w.m.kouw@tudelft.nl

https://arxiv.org/abs/1608.00250