

Reweighting samples under covariate shift using a Wasserstein distance criterion

vendredi 4 décembre 2020 12:00 (30 minutes)

Considering two random variables with different laws to which we only have access through finite size iid samples, we address how to reweight the first sample so that its empirical distribution converges towards the true law of the second sample as the size of both samples goes to infinity. We study an optimal reweighting that minimizes the Wasserstein distance [1] between the empirical measures of the two samples, and leads to an expression of the weights in terms of Nearest Neighbors [2]. The consistency and some asymptotic convergence rates in terms of expected Wasserstein distance are derived, and do not need the assumption of absolute continuity of one random variable with respect to the other. These results have some application in Uncertainty Quantification for decomposition-based estimation [3] and in the bound of the generalization error for the Nearest Neighbor Regression under covariate shift. We then outline the generalization of these results, in which we consider the successive composition of such methods to propagate uncertainty through networks of composite functions.

[1] Villani, Cédric. Optimal transport: old and new. Vol. 338. Springer Science & Business Media, 2008.

[2] Biau, Gérard, and Luc Devroye. Lectures on the nearest neighbor method. Vol. 246. Cham: Springer, 2015.

[3] Amaral, Sergio, Douglas Allaire, and Karen Willcox. "A decomposition-based approach to uncertainty analysis of feed-forward multicomponent systems." International Journal for Numerical Methods in Engineering 100.13 (2014): 982-1005.

Auteurs principaux: TOUBOUL, Adrien (IRT SystemX / Cermics); Dr REYGNER, Julien (Cermics)

Orateur: TOUBOUL, Adrien (IRT SystemX / Cermics)

Classification de Session: Session parallèle 7