

Algorithm

June 1, 2020

1 Algorithm of direct density ratio estimation.

The advantage of estimating the ratio of the probability density function directly from the two samples is that the estimation error can be reduced by estimating the probability density function separately.

Direct estimation of the density ratio uses the unconstrained least squares importance fitting algorithm (uLSIF) [1-3].

First, the density ratio $r(\mathbf{x}) = p(\mathbf{x})/q(\mathbf{x})$ is defined as the sum of kernel functions as

$$\hat{\mathbf{r}}_{\mathbf{w}} = \sum_{i=1}^{N_{kernel}} \mathbf{w}_i \psi_i(\mathbf{x}) = \mathbf{w}^{tr} \psi(\mathbf{x}) \quad (1)$$

, where $\psi(\mathbf{x})$ is non-negative kernel function which is employed gaussian kernel, $p(\mathbf{x})$ and $q(\mathbf{x})$ are probability density function, \mathbf{x} is d-dimensional real number, \mathbf{w} are linear model parameters, N_{kernel} is the number of kernels.

The uLSIF algorithm considers the minimum value of the following function to determine weights \mathbf{w} :

$$E(\mathbf{w}) = \frac{1}{2} \int d\mathbf{x} \{ \hat{r}_{\mathbf{w}}(\mathbf{x}) - r(\mathbf{x}) \}^2 q(\mathbf{x}) \quad (2)$$

$$= \frac{1}{2} \int d\mathbf{x} \mathbf{w}^{tr} \psi(\mathbf{x}) \psi^{tr}(\mathbf{x}) \mathbf{w} q(\mathbf{x}) - \int d\mathbf{x} \mathbf{w}^{tr} \psi(\mathbf{x}) p(\mathbf{x}) + Const. \quad (3)$$

Although the generalized KL divergence is well known as the objective function, the use of KL-div is a convex optimization problem, and the demerit is that the calculation is slow because the optimal solution is obtained by the gradient method. On the other hand, the uLSIF has better computational cost than the gradient method because it can be solved analytically and is theoretically shown that it has excellent properties in terms of convergence and numerical stability of the algorithm [3].

The following equation is obtained by replacing the expected value of the objective function with the sample mean and adding regularization.

$$\min_{\mathbf{w}} \left[\frac{1}{2} \mathbf{w}^{tr} \hat{H} \mathbf{w} - \mathbf{w}^{tr} \hat{h} + \frac{\lambda}{2} |\mathbf{w}|^2 \right] \quad (4)$$

$$\hat{H} = \frac{1}{N_q} \sum_{i=1}^{N_q} \psi(\mathbf{x}_i^q) \psi^{tr}(\mathbf{x}_i^q), \quad \hat{h} = \frac{1}{N_p} \sum_{i=1}^{N_p} \psi(\mathbf{x}_i^p) \quad (5)$$

,where two data samples \mathbf{x}_i^p and \mathbf{x}_i^q are obtained from the pdf $p(\mathbf{x})$ and $q(\mathbf{x})$, respectively, N_p and N_q are the number of two samples.

The analytical \mathbf{w} is obtained as

$$\mathbf{w} = (\hat{H} + \lambda I)^{-1} \hat{h}. \quad (6)$$

If \mathbf{w} is negative during the calculation, it is set to 0.

The bandwidth of RBF σ is set by leave one-out cross validation (LOOCV).

LOOCV can also be analytically obtained.

In the current setup, two sets of samples, $\{x_i^p\}_{i=1}^{N_p}$ and $\{x_j^q\}_{j=1}^{N_q}$, generally have different sample sizes. For $i = 1, \dots, n$, where $n := \min(N_p, N_q)$, suppose that the i -th $p(x)$ sample x_i^p and i -th $q(x)$ sample x_i^q are held out at the same time in the LOOCV procedure.

Let $\hat{r}_i(x)$ be a density-ratio estimate obtained without the i -th $p(x)$ and $q(x)$ samples x_i^p, x_i^q . Then the LOOCV score is expressed as

$$LOOCV = \frac{1}{n_t} \sum_{i=1}^{n_t} \left[\frac{1}{2} (\hat{r}_i(\mathbf{x}_i^q))^2 - \hat{r}_i(\mathbf{x}_i^p) \right]. \quad (7)$$

This package provides a direct density estimation module in which uLSIF [1-3] is implemented. In addition, the relative density ratio (RuLSIF) $p(x)/(\alpha * p(x) + (1 - \alpha) * q(x))$, where α is in the range $[0,1]$, is also implemented [4]. By substituting $q(x)$ into $\alpha * p(x) + (1 - \alpha) * q(x)$, we get RuLSIF algorithm from the above development of formula.

1.0.1 References

- [1] M. Sugiyama, T. Suzuki, and T. Kanamori, **Density Ratio Estimation in Machine Learning**. Cambridge University Press 2012.
- [2] T. Kanamori, S. Hido, and M. Sugiyama, **A least-squares approach to direct importance estimation**, J. Mach. Learn. Res., vol. 10, pp. 1391–1445, 2009.
- [3] T. Kanamori, T. Suzuki, and M. Sugiyama, **Condition number analysis of kernel-based density ratio estimation**, [arXiv](#), (2009).
- [4] M. Yamada, T. Suzuki, T. Kanamori, H. Hachiya, and M. Sugiyama, **Relative density-ratio estimation for robust distribution comparison**, Neural Computation. 2013.