Domain adaptation using optimal transportation

Let $(X_1, Y_1), ..., (X_n, Y_n)$ be an iid training set s.t. $(X_i, Y_i) \backsim \mathbb{P}_{XY}$, with X-marginal $\mathbb{P}$ on $\mathbb{R}^d$, and labels $Y_i \in \{1, ..., K\}$. We are interested in splitting a new data set $\widetilde{X} = (\widetilde{X}_1, ..., \widetilde{X}_M)$, which obeys some unknown modified law $\widetilde{\mathbb{P}}$ into classes $\{1, ..., K\}$.

One can think of $\mathbb{P}$ as a mixture of K distributions $\mathbb{P}_1, ..., \mathbb{P}_K$:

$$X|k \backsim \mathbb{P}_k, \quad \mathbb{P} = \sum_{k=1}^{K} \alpha_k \mathbb{P}_k.$$

The empirical counterparts of $\mathbb{P}_k$ and α_k are then written as follows:

$$\widehat{\mathbb{P}}_k(x) = \frac{1}{r_k} \sum_{i:(x,Y_i),Y_i=k} \delta_x, \quad r_k = \#\{(X_i, Y_i) : \ Y_i = k\}, \quad \widehat{\alpha}_k = \frac{r_k}{N}.$$

We also assume, that there exists an unknown map from $\mathbb{P}_X$ to $\mathbb{P}_{\widetilde{X}}$. The goal is to establish a correspondence between elements of $\widetilde{X}$ and classes $\{1, ..., K\}$ applying the optimal transportation framework. Namely, we are interested in a construction of a *probabilistic classifier* $\mathcal{C}(x)$ on $\widetilde{X}$:

$$\mathcal{C}(x) = (q_1^x, ..., q_K^x),$$

where q_k^x is a probability, that x is classified with k-th label.

Let $\Gamma \in \mathbb{R}_+^{N \times M \times K}$ be a (pseudo)tensor, that describes a meta-transportation plan. Each k th slice $\Gamma_{\cdot,\cdot}^k$ contains a transportation plan between point clouds $\{X_i : \ Y_i = k\}$ and $\{\widetilde{X}_j : \ Y_j = k\}$. We are interested in a partition (classifier Γ^*), that minimize overall transportation costs between all classes:

$$\sum_{k=1}^{K} \left| \sum_{i,j}^{N,M} c_{ij}^k \Gamma_{ij}^k + H\left(r_k, \sum_{i,j}^{N,M} \Gamma_{ij}^k\right) \right| \longrightarrow \inf_{\Gamma},$$

under the following constraints for all $i = \overline{1, N}$, $j = \overline{1, M}$, $k = \overline{1, K}$:

$$\begin{cases} \sum_{ik}^{N,K} \Gamma_{ij}^k = \frac{1}{M}, \\ \sum_{j}^{M} \Gamma_{ij}^k = \widetilde{r}_k \mathbb{P}_k(i), \ \widetilde{r}_k \overset{\text{def}}{=} \sum_{ij}^{N,M} \Gamma_{ij}^k, \\ \Gamma_{ij}^k \geq 0. \end{cases}$$

Here c_{ij}^k is a transportation cost between $X_i|k$ and $\widetilde{X}_j|k$, $H\big(r_k, \sum_{i,j}^{N,M} \Gamma_{ij}^k\big)$ is a penalty for the distance between distribution of mass over classes, i.e. r_k and $\widetilde{r}_k$ should not be too far from each other.

Workpackages:

1. Implement an algorithm that solves OT problem, investigate its theoretical properties;

2. Investigate theoretical properties of the method; given an optimal transportation plan Γ^* construct a classifier for new observations $\widetilde{X}_{\text{new}} \backsim \widetilde{I\!P}$.

Contacts
Statistics: Alexanda Suvorikova, Alexey Kroshnin, Andrey Sobolevskii, Vladimir Spokoiny
Optimisation: Pavel Dvurechensky, VLadimir Spokoiny

Literature

[SAN15] Santambrogio F. Optimal transport for applied mathematicians // Birkaeuser, NY. - 2015.

[CUT14] Cuturi M., Doucet A. Fast computation of Wasserstein barycenters // International Conference on Machine Learning. - 2014. - P. 685-693

[COU17] Courty, Nicolas, et al. "Optimal transport for domain adaptation." IEEE transactions on pattern analysis and machine intelligence 39.9 (2017): 1853-1865.