Inverse Density as an Inverse Problem: the Fredholm Equation Approach

Qichao Que, Mikhail Belkin
Department of Computer Science & Engineering
Ohio State University

Poster ID: Sat55

The Ohio State University

The Problem: Density Ratio Estimation

- **Problem**
  Given samples from two distributions, $x_1^p, \dots, x_n^p \sim p$ and $x_1^q, \dots, x_m^q \sim q$, we want to estimate the density ratio, $\frac{q(x)}{p(x)}$.
- **Our Approach**
  Formulating the problem as a Fredholm integral equation and solving it using kernel methods.
- **A Key Issue**
  Choosing optimal parameters is crucial for performance.

Potential Applications:

- Covariate shift in Transfer Learning, in particular, weighted learning algorithms
- MCMC sampling

Our approach: FIRE Algorithm

By Importance Sampling Equation

$$\mathbb{E}_{x\sim q}(k(x,\cdot)) = \mathbb{E}_{x\sim p}\left(k(x,\cdot)\cdot\frac{q(x)}{p(x)}\right)$$

For any density ρ, we define an Integral Operator $\mathcal{K}_\rho$

$$\mathcal{K}_\rho : h(\cdot) \longmapsto \int k(x,\cdot)h(x)d\rho = \mathbb{E}_{x\sim\rho}(k(x,\cdot)\cdot h(x)).$$

We have an Integral Equation,

$$\mathcal{K}_q 1 = \mathcal{K}_p \frac{q}{p},$$

The Unknown

a.k.a. Fredholm Equation of the first kind. To solve it, we gave the FIRE algorithm, through Regularization in RKHS

$$\frac{q}{p} \approx \arg\min_{h\in\mathcal{H}} \left\| \frac{1}{m}\sum_{i=1}^{m} k(x_i^q,\cdot)\times 1 - \frac{1}{n}\sum_{i=1}^{m} k(x_i^p,\cdot)\times h(x_i^p) \right\|_2^2 + \|f\|_{\mathcal{H}}^2$$

Unsupervised Model Selection and MORE

In our particular setting, with two sets of samples from p and q, Importance Sampling equation gives up the follow model selection method:

- Cross Density Cross Validation (CD-CV)
  Choose test functions $u_1, \ldots, u_l$,
  $$\mathbb{E}_{x\sim q}(u_i(x)) - \mathbb{E}_{x\sim p}\left(u_i(x) \cdot \hat{f}\right) \approx 0,$$
  pick the $\hat{f}$ that achieves the smallest error over all test functions.

Note when $\hat{f} = \frac{q}{p}$, exact equality holds in above.

MORE On the Poster: SAT55

Simple Algorithm!

Theoretically proven convergence!

Unified framework and several variants!