Generative Ratio Matching (GRAM)

Goal

A *stable* learning algorithm for deep generative models with *high* dimensional data

- MMD networks are stable but perform poorly when dimension gets large
- Adversarial methods (GANs, MMD-GANs, etc) are not stable in general

Key ideas

1. Learn a reduced space in which the density ratio between the data and the generator is close to the density ratio in the original space
2. Train the generator via the MMD loss in this reduced space

Matching ratio via minimising squared ratio difference

We'd like to learn a parameterized transformation $f_\theta(x)$ by minimising

$$
\begin{aligned}
D(\theta) &= \int q_x(x) \left(\frac{p_x(x)}{q_x(x)} - \frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))} \right)^2 dx \\
&= C - 2 \int p_x(x) \frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))} dx + \int q_x(x) \left(\frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))} \right)^2 dx \\
&= C - 2 \int \bar{p}(f_\theta(x)) \frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))} df_\theta(x) + \int \bar{q}(f_\theta(x)) \left(\frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))} \right)^2 df_\theta(x) \\
&= C' - \left(\int \bar{q}(f_\theta(x)) \left(\frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))} \right)^2 df_\theta(x) - 1 \right) = C' - \mathrm{PD}(\bar{q}, \bar{p})
\end{aligned}
$$

We can *minimise* the squared ratio difference by *maximising* PD in the reduced space ❤️

Filling up the missing components

- MC estimation of $\mathrm{PD}(\bar{q}, \bar{p}) \approx \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\bar{p}(f_\theta(x_i))}{\bar{q}(f_\theta(x_i))} \right)^2 - 1$ where $x_i^q \sim q_x$
- We only need density ratios $\frac{\bar{p}(f_\theta(x))}{\bar{q}(f_\theta(x))}$ for a set of samples from q during MC.
- We use a MMD based density ratio estimator (Sugiyama et al., 2012) due to its analytical solution under fixed-design setup: $\hat{r}_q = \mathbf{K}_{q,q}^{-1} \mathbf{K}_{q,p} \mathbf{1}$.
  - $\mathbf{K}_{q,q}$ and $\mathbf{K}_{q,p}$ are Gram matrices defined by $[\mathbf{K}_{q,q}]_{i,j} = k(f_\theta(x_i^q), f_\theta(x_j^q))$ and $[\mathbf{K}_{q,p}]_{i,j} = k(f_\theta(x_i^q), f_\theta(x_j^p))$.
- Train the generator via the MMD loss
- Shared Gram matrix between density ratio estimation and generator training
- Simultaneous training of the transform function and the generator

Extra info

Density ratio estimation via (infinite) moment matching

$$\min_{r \in \mathcal{R}} \left\| \int k(x;.)p(x)dx - \int k(x;.)r(x)q(x)dx \right\|_{\mathcal{R}}^2$$

Maximum mean discrepancy

$$\mathrm{MMD}_{\mathcal{F}}(p,q) = \sup_{f \in \mathcal{F}} \left(\mathbb{E}_p[f(x)] - \mathbb{E}_q[f(x)] \right)$$

Gretton et al. (2012) show that it is sufficient to choose $\mathcal{F}$ to be a unit ball in an reproducing kernel Hilbert space $\mathcal{R}$ with a characteristic kernel k. Its MC estimate is

$$\mathrm{M\hat{M}D}_{\mathcal{R}}^2(p,q) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{i'=1}^{N} k(x_i, x_{i'}) - \frac{2}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} k(x_i, y_j) + \frac{1}{M^2} \sum_{j=1}^{M} \sum_{j'=1}^{M} k(y_j, y_{j'})$$