Training generative neural networks via Maximum Mean Discrepancy optimization

Information

`Dziugaite G K, Roy D M, Ghahramani Z. Training generative neural networks via maximum mean d`

Aim

The authors consider training a deep neural network to generate samples from an unknown distribution given i.i.d. data.

Formulate Problem

Given an input Z drawn from some fixed noise distribution $\mathcal{N}$, then to find a function G, called generator, the distribution of the output $G(Z)$ is close to the data's distribution P.

Work

Learning to sample as optimization

It is well known that, for any distribution P and any continuous distribution $\mathcal{N}$ on sufficiently regular space $\mathbb{X}$ and $\mathbb{W}$, respectively, there is a function $G : \mathbb{W} \to \mathbb{X}$, such that $G(W) \sim P$.

For a given family $\{G_\theta\}$ of functions $\mathbb{W} \to \mathbb{X}$, we can cast the problem of learning a generative model as an optimization

$$\arg\min_\theta \delta(P, G_\theta(\mathcal{N}))$$

where δ is some measure of discrepancy. In practice, we only have i.i.d. samples $X_1, X_2, ...$ from P, and so we optimize an empirical estimate of $\delta(P, G\_\{\theta\}(\mathcal{N}))$.

Maximum Mean Discrepancy (MMD)

The MMD between P and $G_\theta(\mathcal{N})$ over $\mathcal{H}$, given by

$$\delta_{MMD_\mathcal{H}}(P, G_\theta(\mathcal{N})) = \sup_{f \in \mathcal{F}} E[f(X)] - E[f(Y)]$$

where $X \sim P$ and $Y \sim G_\theta(\mathcal{N})$. Gretton et al. [1] shows that it can be solved in closed form when $\mathcal{H}$ is a reproducing kernel Hilbert space (RKHS).

Assume that $\mathbb{X}$ is a nonempty compact metric space and $\mathcal{F}$ a class of functions $f : \mathbb{X} \to \mathbb{R}$. Let p and q be Borel probability measures on $\mathbb{X}$, and let X and Y be random variables with distribution p and q, respectively. The maximum mean discrepancy (MMD) between p and q is

$$\mathrm{MMD}(\mathcal{F}, p, q) = \sup_{f \in \mathcal{F}} E[f(X)] - E[f(Y)]$$

[1] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholkopf, ¨ and A. Smola. “A Kernel Two-sample Test”. In: J. Mach. Learn. Res. 13 (Mar. 2012), pp. 723–773.