Lecture 21. Generative modeling with Kernels

Our goal in this lecture is to put together our current knowledge of kernel methods towards one of the modern problems of interest in Data Sci. call Generative modeling.

- AI art generators such as DALL-E & Stable diffusion
- Chatbots
- AI voice generators

etc.

Generative modeling is broadly an unsupervised learning task.

Goal: Given a set of samples $\underline{y}_1, \dots, \underline{y}_n \overset{iid}{\sim} \nu$ (target measure ν) generate a new sample $\underline{y}_{n+1} \sim \nu$.

Why is it difficult? generally $\nu \in \mathbb{P}(\mathcal{X})$ where $\mathcal{X}$ is very high dim. (eg. high resolution images) & we only have access to a small set of samples of ν, ie, $\underline{y}_1, \dots, \underline{y}_n$.

Sampling is an old prob. but generative modeling is relatively new:

- Sampling is broadly the problem of generating iid samples from some measure ν. Typically, we assume we know ν or atleast its density upto a constant.

↪ Markov chain Monte Carlo
(Casella & Robats, "Monte Carlo Statistical Methods")

↪ Variational inference
(Blei et al, "Variational Inference: A review for statisticians")

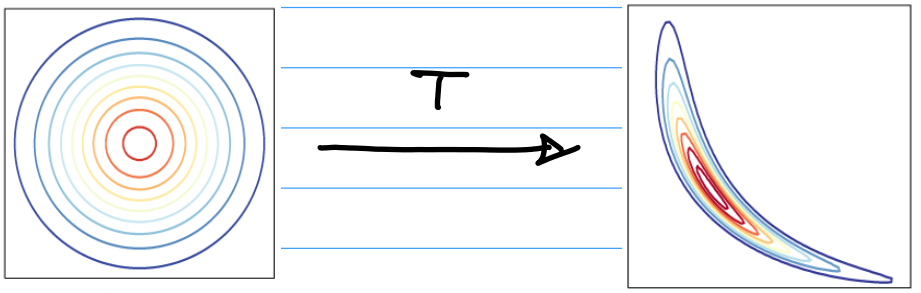
- Generative modeling is a sampling problem
The novelty is that we only have samples from ν!

In this lecture we will cast generative modeling as a trasport problem & propose a solution using kernel methods.

Def^n (Push forward) Given a prob. measure $\mu \in \mathbb{P}(\mathcal{X})$ & a map $T: \mathcal{X} \to \mathcal{Y}$ we define the Push forward of μ with T, denoted as $T_{\#}\mu \in \mathbb{P}(\mathcal{Y})$ as $(T_{\#}\mu)(A) := \mu(T^{-1}(A))$, $\forall$ Borel sets A.
Here $T^{-1}(A) := \{ x \in \mathcal{X} \text{ s.t. } T(x) \in A \}$ (Pre image of A)

Simple intuition for random variables:

If $x \sim \mu$ then $T(x) \sim T_{\#}\mu$.



μ

standard normal

ν

data set "random artist"

Transport is also an old problem & a "hot" math subject these days. It is the main topic in the field of optimal Transport (Villani, "Optimal Transport, old & new").

21.1 Relaxation Using MMD

Problem: Given $Y = \{y_1, \dots, y_n\} \subset \mathbb{R}^d$ where $y_j \overset{iid}{\sim} \nu \in \mathbb{P}(\mathbb{R}^d)$ find a transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that

$$T_{\#}\mu \approx \nu \quad \text{where} \quad \mu = N(0, I).$$

It is natural to formulate the above prob. as an optimization problem for T! that is,

> Find T such that $T_{\#}\mu$ is "close" to $\nu^n = \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}$, what do we mean by close?

(T1)
$$T = \underset{S:\mathbb{R}^d \to \mathbb{R}^d}{\operatorname{argmin}} \; MMD_K(S_{\#}\mu, \nu^n)$$

where $K: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is an appropriate kernel such as RBF or RQ (ϵ lengthscale should be tuned!).

Dealing with $T_{\#}\mu$ is a bit cumbersome so let us replace μ also with an empirical measure.

$$\mu^n = \frac{1}{n}\sum_{j=1}^{n} \delta_{x_j}, \qquad x_j \sim \mu = N(0, I)$$

$$\Rightarrow T_{\#}\mu \approx T_{\#}\mu^n = \frac{1}{n}\sum_{j=1}^{n} \delta_{T(x_j)}$$

Notation: $Y = \{y_1, \ldots, y_n\}$

$X = \{x_1, \ldots, x_n\}$

$T(X) = \{T(x_1), \ldots, T(x_n)\}$

Farther approx. (T1) with

(T2) $$T = \underset{S:\mathbb{R}^d \to \mathbb{R}^d}{\operatorname{argmin}} \; MMD_k(S(X), Y)$$

Observe, this is now a nice loss since we can easily evaluate it for any choice of S!

But, the map T can be anything! in fact, prob. has. infinitely many solutions!

21.2 Regularization with RKHS norms & a Rep. formula

We would like to bring some stability to the problem by regularizing T. Also, in practice, having a T that is "nicer / more regular" would be beneficial. So, we may decide to penalize on RKHS norm of T.

Challenge: $T: \mathbb{R}^d \to \mathbb{R}^d$ how do we make sense of its RKHS norm?

Simple solution: Simply add up the RKHS norms of components of the map

$$T(\underline{x}) = \left(T^1(\underline{x}), T^2(\underline{x}), \ldots, T^d(\underline{x})\right)^T \in \mathbb{R}^d$$

$$\|T\|_Q^2 := \sum_{j=1}^{d} \|T^j\|_Q^2$$

for some kernel $Q : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, eg. RBF, Matérn, Polynomial. etc.

We can now define our regularized transport problem:

$$\text{(T3)} \quad T = \operatorname*{argmin}_{S:\mathbb{R}^d \to \mathbb{R}^d} MMD_K(S(X), Y) + \lambda \sum_{j=1}^{d} \|S^j\|_Q^2$$

WARNING: The kernels K & Q are not necessarily related! They need to be chosen & tuned separately

Observe, prob. (T3) can also be viewed as a "regression" prob. for the components T^j of T! So we can write a rep. formula for it!

Let us write $S(x_i) = z_i$, $i = 1, \dots, n$ & $Z = S(X)$ then we can write (T3) as the new prob.

$$\min_{z=\{z_1,\dots,z_n\}} \quad \min_{S:\mathbb{R}^d \to \mathbb{R}^d} \quad MMD_k(z,Y) + \lambda \sum_{j=1}^{d} \|S^j\|_Q^2$$

$$\text{s.t. } S(x_i) = z_i \;,\; i=1,\dots,n$$

equivalently ↳

$$S^j(x_i) = z_{ij}, \quad \begin{matrix} i=1,\dots,n \\ j=1,\dots,d \end{matrix}$$

Minimizing over S for a fixed z we get that

$$S^j = Q(\cdot, X)\, Q(X,X)^{-1} (z_{:j})^T$$

& furthermore $\|S^j\|_Q^2 = z_{:j}\, Q(X,X)^{-1} (z_{:j})^T$

so the rep. thm takes the form

$$T(x) = (T^1(x), \dots, T^d(x))^T$$

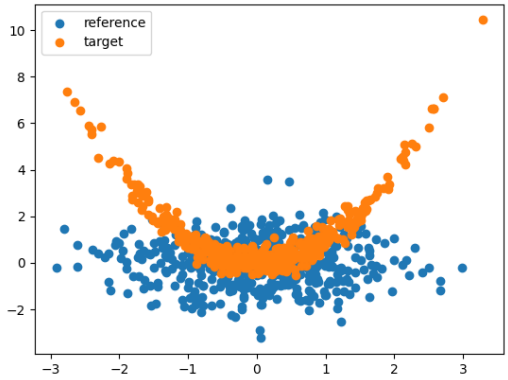
$$T^j = Q(\cdot, X)\, Q(X,X)^{-1} (z_{:j})^T, \quad j=1,\dots,d$$

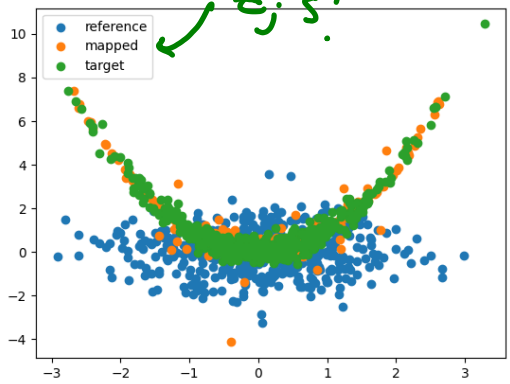
where $z \in \mathbb{R}^{n \times d}$ solves

$$\min_z \; MMD_k(z,Y) + \lambda \sum_{j=1}^{d} z_{:j}\, Q(X,X)^{-1} (z_{:j})^T$$

The above Prob. is easily solvable using off-the shelf optimizers.

mapping a 2D Gaussian to a non gaussian measure





2 g's!

The code uses a RF implementation of MMD for speed. Also looks for maps of the form

$\underline{x} \longmapsto \underline{x} + T(\underline{x})$ rather than $\underline{x} \longmapsto T(\underline{x})$