

# 1 Query Release via Moment Matching

In this section, we show how we can use moment-matching kernels as an efficient distance metric between distributions. We can choose a kernel such that the feature map contains all queries of interest, and the kernel trick allows us to compute the worst expected difference between two distributions in polynomial time.

## 1.1 Moment Matching Kernels

### 1.1.1 Reproducing Kernel Hilbert Spaces

**Definition 1:** Let  $\mathcal{F}$  be a class of functions and let  $p$  and  $q$  be the true data and fake data distribution respectively, and  $X$  and  $Y$  be finite observations drawn iid from  $p, q$  respectively. We then define the maximum mean discrepancy and its empirical estimate as

$$MMD[\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} (\mathbb{E}_{x \sim p}[f(x)] - \mathbb{E}_{y \sim q}[f(y)]) \quad (1)$$

$$MMD[\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left( \frac{1}{m} \sum_{i=1}^m f(x_i) - \frac{1}{N} \sum_{i=1}^m f(y_i) \right) \quad (2)$$

We now let  $\mathcal{F}$  be the unit ball in some reproducing kernel Hilbert space  $\mathcal{H}$ . By the representer theorem, there is a feature map  $\phi(x)$  from  $\mathcal{X}$  to  $\mathbb{R}$  such that  $f(x) = \langle f, \phi(x) \rangle_{\mathcal{H}}$ . In canonical form, this feature map is  $\phi(x) = k(x, \cdot)$ . In this instance, we can express the MMD as the distance in  $\mathcal{H}$  between mean embeddings [1].

$$MMD^2[\mathcal{F}, p, q] = \|\mathbb{E}_{x \sim p}[f(x)] - \mathbb{E}_{y \sim q}[f(y)]\|_{\mathcal{H}}^2 \quad (3)$$

### 1.1.2 Query Classes as feature maps

Using the kernel trick, we can efficiently find the function  $f \in \mathcal{F}$  s.t  $\|f\|_{\mathcal{H}} \leq 1$  that maximizes the difference :

$$\|\mathbb{E}_{x \sim p}[f(x)] - \mathbb{E}_{y \sim q}[f(y)]\|_{\mathcal{H}}$$

[TODO can't make the jump from supremum to sum]

## 1.2 Boolean Kernels

Note that maximizing this discrepancy is exactly equivalent to maximizing the Wasserstein GAN objective. As such, if we can find a kernel  $k$  in a RKHS such that its feature map  $\mathcal{F}$  describes a query class of interest, we can use  $\mathcal{L}_{MMD^2}$  as an efficient loss function for the *entire* class of queries.

One such kernel is the kernel corresponding to all monotone monomials of length up to  $d$ , which we denote by  $k_d$  [2]. [TODO define  $\mathbb{B}$ ]

$$k_d(\mathbf{x}, \mathbf{x}') := \langle f(\mathbf{x}), f(\mathbf{x}') \rangle_K = \sum_{\mathbf{i} \in \mathbb{B}_d^N} K_{\|\mathbf{i}\|}^{-1} \mathbf{x}^{\mathbf{i}} \mathbf{x}'^{\mathbf{i}}$$

Explicitly computing this would require summing  $|\mathbb{B}_d^n| = O(n^d)$  terms.  
 [TODO explain why, prove this], Thus,

$$k(\mathbf{x}, \mathbf{x}') = \sum_{j=0}^d e^j \binom{\langle \mathbf{x}, \mathbf{x}' \rangle}{j} = (1 + b)^{\langle \mathbf{x}, \mathbf{x}' \rangle}$$

where  $b$  determines the weight assigned to higher order polynomials. a  
 [TODO all subsets kernel]

**Note:** A supremum of convex functions is convex

## References

- [1] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander J. Smola. A Kernel Two-Sample Test. *J. Mach. Learn. Res.*, 13:723–773, 2012.
- [2] Mirko Polato, Ivano Lauriola, and Fabio Aiolli. A Novel Boolean Kernels Family for Categorical Data. *Entropy*, 20(6):444, June 2018.