

随机傅里叶特征(Random Fourier Features)

作者: 凯鲁嘎吉 - 博客园 <http://www.cnblogs.com/kailugaji/>

1. 基本知识

1.1 傅里叶变换与逆傅里叶变换

We may think of a non-periodic signal as one arising from a periodic signal in which the period extends from $-\infty$ to $+\infty$. Then, for a signal that is a function of time with period from $-\infty$ to $+\infty$, we form the integral

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t} dt \quad (8.1)$$

and assuming that it exists for every value of the radian frequency ω , we call the function $F(\omega)$ the Fourier transform or the Fourier integral.

The Fourier transform is, in general, a complex function. We can express it as the sum of its real and imaginary components, or in exponential form, that is, as

$$F(\omega) = \text{Re}\{F(\omega)\} + j\text{Im}\{F(\omega)\} = |F(\omega)|e^{j\varphi(\omega)} \quad (8.2)$$

The Inverse Fourier transform is defined as

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega)e^{j\omega t} d\omega \quad (8.3)$$

1.2 平移不变性核

A kernel function $k : R^D \times R^D \rightarrow R$ is so-called shift-variant if $k(x_i, x_j) = k(x_i - x_j)$

具有平移不变性的核有：

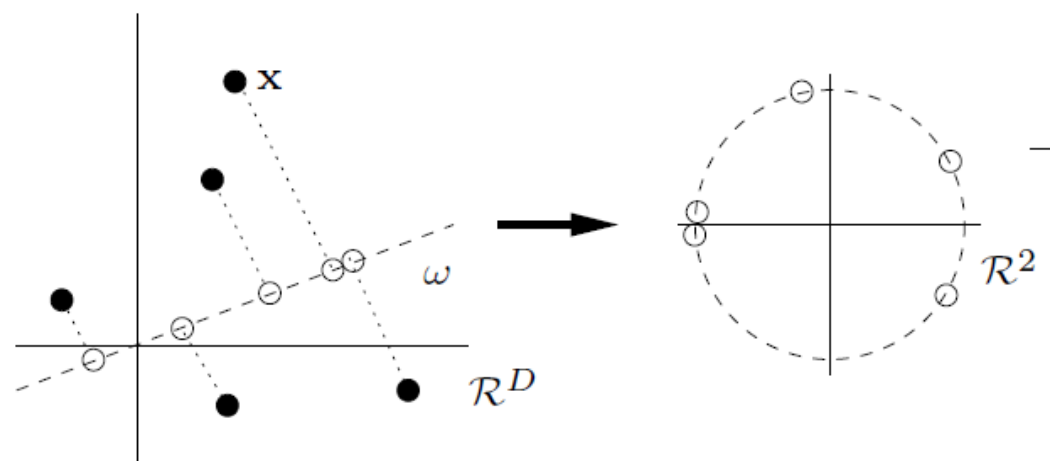
Kernel Name	$k(\Delta)$	$p(\omega)$
Gaussian	$e^{-\frac{\ \Delta\ _2^2}{2}}$	$(2\pi)^{-\frac{D}{2}} e^{-\frac{\ \omega\ _2^2}{2}}$
Laplacian	$e^{-\ \Delta\ _1}$	$\prod_d \frac{1}{\pi(1+\omega_d^2)}$
Cauchy	$\prod_d \frac{2}{1+\Delta_d^2}$	$e^{-\ \Delta\ _1}$

1.3 Bochner定理

If a shift-variant function is a kernel function, there exists a non-negative measure $p(\omega)$ such that the function $k(\delta)$ is the Fourier transformation of $p(\omega)$.

2. 介绍

随机傅里叶特征方法是一种近似核函数的方法，旨在找到一个低阶的映射函数 $z(x)$ ，将D维原始数据映射到d维随机特征空间中，随机特征空间中两个特征的内积约等于核函数，即 $z(x)'z(y) \approx k(x,y) = k(x-y)$



算法具体流程：

Algorithm 1 Random Fourier Features.

Require: A positive definite shift-invariant kernel $k(\mathbf{x}, \mathbf{y}) = k(\mathbf{x} - \mathbf{y})$.

Ensure: A randomized feature map $\mathbf{z}(\mathbf{x}) : \mathcal{R}^d \rightarrow \mathcal{R}^D$ so that $\mathbf{z}(\mathbf{x})' \mathbf{z}(\mathbf{y}) \approx k(\mathbf{x} - \mathbf{y})$.

Compute the Fourier transform p of the kernel k : $p(\omega) = \frac{1}{2\pi} \int e^{-j\omega' \delta} k(\delta) d\Delta$.

Draw D iid samples $\omega_1, \dots, \omega_D \in \mathcal{R}^d$ from p and D iid samples $b_1, \dots, b_D \in \mathcal{R}$ from the uniform distribution on $[0, 2\pi]$.

Let $\mathbf{z}(\mathbf{x}) \equiv \sqrt{\frac{2}{D}} [\cos(\omega_1' \mathbf{x} + b_1) \dots \cos(\omega_D' \mathbf{x} + b_D)]'$.

注：我的表述与算法流程中的不一样，我的表述是 \mathbf{z} ：从 $D \rightarrow d$ 维。

3. 具体推导

给定一个平移不变性的核，怎样找到这样的映射 $\mathbf{z}(\mathbf{x})$ 呢？

$$\begin{aligned}
k(x_i, x_j) &\stackrel{(1)}{=} k(x_i - x_j) \stackrel{(2)}{=} \int_{\mathbb{R}^d} p(w) e^{-i w^T (x_i - x_j)} dw \stackrel{(3)}{=} \mathbf{E}_w \left[e^{-i w^T (x_i - x_j)} \right] \\
&\stackrel{(4)}{=} \mathbf{E}_w \left[\cos(w^T (x_i - x_j)) \right] \stackrel{(5)}{=} \mathbf{E}_w \left[z_w(x_i)^T z_w(x_j) \right] \stackrel{(6)}{\approx} \frac{1}{d} \sum_{m=1}^d z_{w_m}(x_i) z_{w_m}(x_j) = z(x_i)^T z(x_j)
\end{aligned}$$

其中,

$$z_{w_m}(x) = \sqrt{2} \cos(w_m^T x + b_m), w_m \sim p(w), b_m \sim U[0, 2\pi]$$

(1) 平移不变性

(2) 傅里叶变换与逆傅里叶变换

(3) $w \sim p(w)$

(4) $e^{-i\theta} = \cos(\theta) - i \sin(\theta)$, 只取实部

(5) 具体推导过程

➤ 具体推导过程

$$\begin{aligned}
 \mathbf{E}_w [z_w(x_i)^T z_w(x_j)] &= \mathbf{E}_w [\sqrt{2} \cos(w^T x_i + b) \sqrt{2} \cos(w^T x_j + b)] \\
 &= \mathbf{E}_w [\cos(w^T (x_i + x_j) + 2b)] + \mathbf{E}_w [\cos(w^T (x_i - x_j))] \\
 &= \mathbf{E}_w [\cos(w^T (x_i - x_j))]
 \end{aligned}$$

$t = \omega^T (x - y)$ and let $2b = x$. Therefore, $x \sim \text{Uniform}(0, \pi)$ and $p(x) = 1/\pi$. Then

$$\begin{aligned}
 2\cos(x)\cos(y) &= \cos(x+y) + \cos(x-y) \\
 \mathbf{E}_w [\cos(w^T (x_i + x_j) + 2b)] &= \mathbf{E}_w [\mathbf{E}_b [\cos(w^T (x_i + x_j) + 2b) | w]] \\
 w &\sim p(w), b \sim U[0, 2\pi]
 \end{aligned}$$

0

$$\begin{aligned}
 \mathbf{E}[\cos(t+x)] &= \int_0^\pi \frac{\cos(t+x)}{\pi} dx \\
 &= \frac{1}{\pi} \int_0^\pi \cos(t+x) dx \\
 &= \frac{1}{\pi} [\sin(t+x)]_0^\pi \\
 &= \frac{1}{\pi} [\sin(t) - \sin(t+\pi)] \\
 &= 0
 \end{aligned}$$

The last step holds because $\sin(t) = \sin(t \pm \pi)$.

(6) 蒙特卡洛方法

4. 结论

至此，找到随机映射 $z(x)$ ，使得原始数据 x 映射到 d 维随机特征空间中。

$$z(x) = \frac{1}{\sqrt{d}} \left(\sqrt{2} \cos(w_1^T x + b_1), \dots, \sqrt{2} \cos(w_d^T x + b_d) \right)^T$$

我的理解：

原先支持向量机(SVM)通过核函数将样本从原始空间映射到一个更高维的特征空间，使得样本在这个特征空间内线性可分，而问题是这个映射函数没法显式地求出具体表达式，这个特征空间维度不确定，可能很高，甚至是无穷维，为了避免计算映射函数，于是用两个样本在特征空间的内积等于它们在原始样本空间通过函数 $k(x_i, x_j)$ 计算的结果来进行计算，这样就不用直接去计算高维甚至无穷维特征空间中的内积。但是当训练样本个数很大时，用SVM求解时间复杂度很高。

现在是找一个映射函数 z 来近似原先的映射函数，找到的映射函数 z 有显式的表达式， z 的维度是可以人为控制的，是已知的，这个映射函数与原先的SVM的映射函数相比，维度要低，并且在新的特征空间中训练线性学习器，计算复杂度也比原先的低，在大规模数据集上可以有效提高计算效率。但是随机特征空间维度 d 的选择是一个值得思考的问题，实验表明，随着随机特征维度的增加，结果越好，这是因为高维度的随机特征能够更好地近似核函数，但维度升高也会导致时间复杂度增加，因此如何在保证结果较好的情况下，又不会使时间复杂度很高，如何选择 d ，能对我们的结果有利，是一个值得研究的问题。

文献截图中的- j 就是我这里的 i 。

5. 参考文献

[1] Rahimi and B. Recht, "[Random features for large-scale kernel machines](#)," in Advances in Neural Information Processing Systems 20, pp. 1177-1184, Curran Associates, Inc., 2008.

代码: <https://github.com/kma32527/Random-Fourier-Features>

[2] Random Fourier Features <http://gregorygundersen.com/blog/2019/12/23/random-fourier-features>

[3] Karris S T. Signals and systems with MATLAB computing and Simulink modeling[M]. Orchard publications, 2007.

[4] 王迎旭. 基于随机特征的多核分布式协同模糊聚类算法研究[D]. 济南大学, 2019.