1. Kernels and SPD functions

(a) $k(x,y) = \cos(x-y)$, $\mathbb{R}\times\mathbb{R}$

$k(x,y) = k(y,x)$, $\forall x,y \in \mathbb{R}$. Take $x_1, \cdots x_n \in \mathbb{R}$ $\alpha_1 \cdots \alpha_n \in \mathbb{R}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i\alpha_j \cos x_i \cos x_j + \alpha_i\alpha_j \sin x_i \sin x_j)$$

$$= \left(\sum_{i=1}^{n}\alpha_i \cos x_i\right)^2 + \left(\sum_{i=1}^{n}\alpha_i \sin x_i\right)^2 \geq 0 \quad \#$$

(b) $k(x,y) = \cos(x^2-y^2)$, $\mathbb{R}\times\mathbb{R}$

$k(x,y) = k(y,x)$, $\forall x,y \in \mathbb{R}$. Take $x_1 \cdots x_n \in \mathbb{R}$, $\alpha_1 \cdots \alpha_n \in \mathbb{R}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j k(x_i, x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j \left[\cos(x_i^2)\cos(y_i^2) + \sin(x_i^2)\sin(y_i^2)\right]$$

$$= \left[\sum_{i=1}^{n}\alpha_i \cos(x_i^2)\right]^2 + \left[\sum_{i=1}^{n}\alpha_i \sin(x_i^2)\right]^2 \geq 0 \quad \#$$

(c) $k(x,y) = \frac{1}{x+y}$, $\mathbb{R}_+ \times \mathbb{R}_+$

$k(x,y) = k(y,x)$, $\forall x,y \in \mathbb{R}_+$ Take $x_1 \cdots x_n \in \mathbb{R}_+$ $\alpha_1 \cdots \alpha_n \in \mathbb{R}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\alpha_i\alpha_j}{x_i+x_j} = \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j \int_1^\infty t^{-1-x_i-x_j}\,dt = \int_1^\infty t^{-1}\sum_{i=1}^{n}\frac{\alpha_i}{t^{x_i}}\sum_{j=1}^{n}\frac{\alpha_j}{t^{x_j}}\,dt$$

$$= \int_1^\infty t^{-1}\left(\sum_{i=1}^{n}\frac{\alpha_i}{t^{x_i}}\right)^2 dt \geq 0 \quad \#$$

(d) $k(x,y) = e^{-\|x-y\|_1}$, $\mathbb{R}^d \times \mathbb{R}^d$

$k(x,y) = k(y,x)$, $\forall x,y \in \mathbb{R}^d$. Take $x_1 \cdots x_n \in \mathbb{R}^d$, $\alpha_1 \cdots \alpha_n \in \mathbb{R}$

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j k(x_i, x_j) \overset{(*)}{=} \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j \int_{\mathbb{R}^d}\prod_{k=1}^{d}\frac{1}{\pi(1+w_k^2)} e^{iw^T(x_i-x_j)}\,dw$$

$$= \int_{\mathbb{R}^d}\prod_{k=1}^{d}\frac{1}{\pi(1+w_k^2)}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{iw^T x_i}\,\overline{e^{iw^T x_j}}\,\alpha_i\alpha_j\,dw$$

$$= \int_{\mathbb{R}^d}\prod_{k=1}^{d}\frac{1}{\pi(1+w_k^2)}\left\|\sum_{i=1}^{n}\alpha_i e^{iw^T x_i}\right\|^2 dw \geq 0,$$ where $\|a+bi\| = \sqrt{a^2+b^2}$, $a, b \in \mathbb{R}$.

and $(*)$ is due to Fourier transform. #

2 Error analysis of random feature model

$k: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ $k(x,y) = \mathbb{E}_{w\sim\pi} [\varphi(x, w) \varphi(y, w)]$

$\varphi: \mathcal{X} \times \Omega \to \mathbb{R}$

$$\hat{k}_m(x,y) = \frac{1}{m}\sum_{j=1}^{m} \varphi(x, w_j) \varphi(y, w_j), \quad w_j \overset{iid}{\sim} \pi$$

$W = (w_1 \cdots w_m)$ $\quad \mathcal{E}_m(W) = \| \hat{k}_m - k \|_{L^2} = \sqrt{\mathbb{E}_{x,y} | k(x,y) - \hat{k}_m(x,y) |^2}$ - L_2 error

$Q = \mathbb{E}_{x,y,w} [\varphi^2(x;w) \varphi^2(y;w)]$

$$\mathbb{E}_W [\mathcal{E}_m(W)] \le \sqrt{\frac{Q}{m}}.$$

Proof. By Cauchy-Schwarz inequality, we have

$$\left[\mathbb{E}_{w_1 \cdots w_m} [\mathcal{E}_m(W)] \right]^2 \le \mathbb{E}_{w_1 \cdots w_m} \mathcal{E}_m^2(W) \, \mathbb{E}_{w_1 \cdots w_m} 1$$

Thus $LHS^2 = \left(\mathbb{E}_{w_1 \cdots w_m} [\mathcal{E}_m(W)]\right)^2 \le \mathbb{E}_{w_1 \cdots w_m} \mathcal{E}_m^2(W)$

$$= \mathbb{E}_{w_1 \cdots w_m} \mathbb{E}_{x,y} | k(x,y) - \hat{k}_m(x,y) |^2$$

$$= \mathbb{E}_{x,y} \mathbb{E}_{w_1 \cdots w_m} | k(x,y) - \hat{k}_m(x,y) |^2$$

$$= \frac{1}{m^2} \mathbb{E}_{x,y} \mathbb{E}_{w_1 \cdots w_m} \left| \sum_{j=1}^{m} \varphi(x, w_j) \varphi(y; w_j) - m \mathbb{E}_w \varphi(x;w) \varphi(y;w) \right|^2$$

Fix x,y, denote by $Y_j(w_j) = \varphi(x; w_j) \varphi(y; w_j)$

$LHS^2 \le \frac{1}{m^2} \mathbb{E}_{x,y} \mathrm{Cov}_{w_1 \cdots w_m} \left(\sum_{j=1}^{m} Y_j \right)$ (definition of covariance)

$= \frac{1}{m} \mathbb{E}_{x,y} \mathrm{Cov}_w Y_1$ (because $w_1 \cdots w_n \overset{iid}{\sim} w$

$= \frac{1}{m} \mathbb{E}_{x,y} [\mathbb{E} Y_1^2 - (\mathbb{E} Y_1)^2]$

$\le \frac{1}{m} \mathbb{E}_{x,y} \mathbb{E} Y_1^2$

$= \frac{1}{m} \mathbb{E}_{x,y} \mathbb{E}_{w\sim\pi} \varphi^2(x;w) \varphi^2(y;w)$

$= \frac{1}{m} Q$

Thus $LHS \le \sqrt{\frac{Q}{m}}$ # $\quad$ where LHS denotes 'left-hand side'

3 LogSumExp trick

$z_1, \dots, z_n \in \mathbb{R}$. $\beta > 0$,

$$\max_j z_j \le \frac{1}{\beta} \log \sum_{j=1}^{n} e^{\beta z_j} \le \max_j z_j + \frac{\log(n)}{\beta}$$

Proof. Let $j_0 = \arg\max_j z_j = \arg\max_j \beta z_j = \arg\max_j e^{\beta z_j}$

$$\frac{1}{\beta} \log \sum_{j=1}^{n} e^{\beta z_j} \ge \frac{1}{\beta} \log e^{\beta z_{j_0}} = \frac{1}{\beta} \beta z_{j_0} = z_{j_0} = \max_j z_j$$

$$\frac{1}{\beta} \log \sum_{j=1}^{n} e^{\beta z_j} \le \frac{1}{\beta} \log \sum_{j=1}^{n} e^{\beta z_{j_0}} = \frac{1}{\beta} \log (n e^{\beta z_{j_0}})$$

$$= \frac{\log(n)}{\beta} + z_{j_0} = \frac{\log(n)}{\beta} + \max_j z_j$$

#