Random Fourier features

A kernel $k(x,y)$ is shift-invariant if

$$k(x,y) = k(x-y)$$

Bochner's Theorem

A shift-invariant function $k(x,y) = k(x-y)$ is a valid kernel iff $k(\cdot)$ is the Fourier tranform of a non-negative measure.

$$k(\Delta) = \int \exp(j\langle \omega, \Delta\rangle)\, \mu(\omega)\, d\omega$$

$$\text{or } k(\Delta) = Z \cdot \underset{\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim\sim}{E_{\omega\sim p(\omega)}[\exp(j\langle \omega, \Delta\rangle)]}$$

where $Z = \int \mu(\omega)\, d\omega$, $p(\omega) = \frac{\mu(\omega)}{Z}$

Real-valued features

$\underline{k(x,y) \text{ is real-valued}}$, we can identify the following real-valued feature map:

$$\Phi_{\omega,1}(x) := \sqrt{\frac{Z}{K}} \Big(\cos(\langle \omega_1, x\rangle), \cdots, \cos(\langle \omega_K, x\rangle),$$
$$\sin(\langle \omega_1, x\rangle), \cdots, \sin(\langle \omega_K, x\rangle) \Big) \in \mathbb{R}^{2K}$$

$\omega_1, \cdots, \omega_K$ i.i.d. drawn from $p(\omega)$

$$\begin{aligned}
k(x,y) &= Re(k(x,y)) \\
&= Z \cdot E_{\omega\sim p(\omega)}[Re(\exp(j\langle \omega, x-y\rangle))] \\
&= Z \cdot E_{\omega\sim p(\omega)}[\cos(\langle \omega, x-y\rangle)] \\
&= \underline{Z \cdot E_{\omega\sim p(\omega)}}[\cos(\langle \omega, x\rangle - \langle \omega, y\rangle)]
\end{aligned}$$

W.T.S: $k(x,y) = \underline{Z \cdot \mathbb{E}_{\substack{w \sim p(w), \\ b \sim Unif[0,2\pi)}}} \left[2\cos(\langle w, x \rangle + b) \cdot \cos(\langle w, y \rangle + b) \right]$

Sample i.i.d. $\{(w_i, b_i)\}_{i=1}^{K}$ with $w_i \sim p(w)$
$b_i \sim Unif[0, 2\pi]$

$\Rightarrow \Phi_{w,b}(x) = \sqrt{\frac{Z}{K}} \left(\cos(\langle w_1, x \rangle + b_1), \cdots, \cos(\langle w_K, x \rangle, b_K) \right) \in \mathbb{R}^K$

$\rightarrow$ $\Phi_{w,b}(x)$ is named "Cosine Random Fourier Feature"

The "random phase" $\{b_i\}_{i=1}^{K}$ compress two coordinates into one cosine with a phase

W.T.S: $\mathbb{E}_{b \sim Unif[0,2\pi]} \left[2\cos(\underbrace{\langle w, x \rangle}_{:=u} + b) \cdot \cos(\underbrace{\langle w, y \rangle}_{:=v} + b) \right]$
$= \cos(\langle w, x \rangle - \langle w, y \rangle)$

Trig Expansion:

$2\cos x \cos y = \cos(x-y) + \cos(x+y)$

$\mathbb{E}_{b \sim Unif}\left[2\cos(u+b)\cos(v+b) \right] = \mathbb{E}\left[\cos(u-v) + \cos(u+v+2b) \right]$

$= \frac{1}{2\pi} \int_0^{2\pi} \left[\cos(u-v) + \underbrace{\cos(u+v+2b)}_{=0} \right] db$

$= \cos(u-v) = \cos(\langle w, x \rangle - \langle w, y \rangle)$ $\boxtimes$

$k(x,y) = \underbrace{1}_{\text{scalar. hence a trivial kernel}} + \min(x,y)$

WTS: $\min(x,y)$ is a valid kernel.

$$\min(x,y) = \int_0^1 \underbrace{\mathbb{1}\{t \le x\}\,\mathbb{1}\{t \le y\}}_{:= \mathbb{1}\{t \le \min(x,y)\}}\, dt$$

$$= \int_0^{\min(x,y)} 1 \, dt$$

We found the feature map of $\min(x,y)$, so $\min(x,y)$ is a valid kernel.

By addition rule, $k(x,y)$ is also a kernel

Kernelized k-means

$x_1, \cdots, x_n \in \mathbb{R}^d$

$\downarrow$

clusters $C_1, \cdots, C_k$

minimize the distortion

$$J(C_1, \cdots, C_k) = \sum_{r=1}^{k} \sum_{i \in [|C_r|]} \|x_i - \mu_r\|^2$$

where $\mu_r = \frac{1}{|C_r|} \sum_{j \in [|C_r|]} x_j$

Step 1: Rewrite in terms of inner products

$$\|x_i - \mu_r\|^2 = \langle x_i, x_i \rangle - \frac{2}{|C_r|} \sum_{j \in [|C_r|]} \langle x_i, x_j \rangle + \frac{1}{|C_r|^2} \sum_{j,l \in [|C_r|]} \langle x_j, x_l \rangle$$

(The objective depends only on inner products)

Step2: Kernelization

Replace $\langle x_i, x_j \rangle$ by $k(x_i, x_j)$

Step3: Algorithm

① Compute Gram matrix

$$K = (k(x_i, x_j))_{i,j=1}^{n}$$

② Initialize k clusters (k-means++)

③ Assignment

For each x_i and C_j

$$\underset{j}{\arg\min}\, D(x_i, C_j) =$$

$$= k(x_i, x_j) - \frac{2}{|C_j|} \sum_{j \in [|C_j|]} k(x_i, x_j) + \frac{1}{|C_r|^2} \sum_{j,l \in [|C_j|]} k(x_j, x_l)$$

④ Update step

Recompute $|C_j|$ & sum of kernel values

⑤ Repeat steps ③ - ④ until assignments converge