CSE 546 Homework 4B

Jake Gonzales

December 11, 2023

Collaborators: n/a

1 Random Fourier Features

B1.

(a) In this sub-problem, we'll use the mentioned Fourier-transform interpretation of k to derive a randomized mapping $z : \mathbb{R}^d \mapsto \mathbb{R}^D$ which is an unbiased estimate of the kernel function, i.e.

$$\mathbb{E}_w[z(x)^T z(x')] = k_p(x, x').$$

If $z(x)^T z(x')$ serves as a good approximation to the kernel matrix, we can apply the aforementioned approximation algorithm.

i) By Bochner's theorem we have that a continuous kernel $k(x, x') = k(x - x')$ on $\mathbb{R}^n$ is positive definite if and only if k is the Fourier transform of a non-negative measure. Therefore we can express the kernel as follows: for a probability distribution $p(w)$ define

$$k_p(x, x') = k(x - x') = \int_{R^d} p(w) e^{iw^T(x-x')} dw = \mathbb{E}_w[e^{iw^T(x-x')}] \tag{1}$$

Define Euler's formula $e^{iy} = \cos(y) + i\sin(y)$. Since both our distribution and kernel are real-valued, we can write

$$\begin{aligned} e^{iw^T(x-x')} &= \cos(w^T(x - x')) - i\sin(w^T(x - x')) \\ &= \cos(w^T(x - x')). \end{aligned}$$

Therefore using Euler's formula and plugging into (1) we have shown that

$$k_p(x, x') = \mathbb{E}_w[\cos(w^T(x - x'))].$$

ii) We define $z_w : \mathbb{R}^d \mapsto \mathbb{R}$ as

$$z_w(x) = \sqrt(2)\cos(w^T x + b)$$

where $w \sim p(w), b \sim \text{Uniform}(0, 2\pi)$. We can show that the expected product of $z_w(x)$ is an unbiased estimate of the kernel function by

$$\begin{aligned} \mathbb{E}_{w,b}[z_w(x) z_w(x')] &= \mathbb{E}_{w,b}[\sqrt(2)\cos(w^T x + b)\sqrt(2)\cos(w^T x' + b)] \\ &= \mathbb{E}_{w,b}[\cos(w^T(x + x') + 2b)] + \mathbb{E}_{w,b}[\cos(w^T(x - x'))] \\ &= \mathbb{E}_{w,b}[\cos(w^T(x - x'))] \\ &= k_p(x, x'). \end{aligned}$$

where the second equivalence comes by using the following trigonometric identity: $2\cos(\alpha)\cos(\beta) = \cos(\alpha + \beta) + \cos(\alpha - \beta)$ and the third equivalence uses the fact that the expectation with respect to b is zero.

iii) Given z as the d-dimensional concatenation of $z_w(x)$ as

$$z(x) = \left[\frac{1}{\sqrt{D}} z_{w_1}(x), \frac{1}{\sqrt{D}} z_{w_2}(x), \cdots, \frac{1}{\sqrt{D}} z_{w_D}(x)\right]$$

then we can write

$$\begin{aligned}
\mathbb{E}_w[z(x)^T z(x')] &= \frac{1}{D}\sum_{d=1}^{D} z_{w_d}(x) z_{w_d}(x') \\
&= \frac{1}{D}\sum_{d=1}^{D} 2\text{cos}(w_d^T x + b_d)\text{cos}(w_d^T x' + b_d) \\
&= \frac{1}{D}\sum_{d=1}^{D} \text{cos}(w_d^T (x - x')) \\
&\approx \mathbb{E}_w[\text{cos}(w^T(x - x'))] \\
&= k_p(x, x').
\end{aligned}$$

(b) n/a

(c) Hoeffding's inequality provides an upper bound on the probability that the sum of bounded independent random variables deviates from its expected value by more than a certain amount. In part a it is shown that the estimate of the kernel is formed by concatenating D randomly chosen z_w into the D-dimensional vector $z(x)$ and normalizing each component by $\sqrt{D}$. Since z_w is bounded between $+\sqrt{2}$ and $-\sqrt{2}$ for a fixed pair of points x and x' then we can use Hoeffding's inequality to provide a guarentee for exponentially fast convergence in D between $z(x)z(x')$ and $k(x, x')$.