Random Kitchen Sinks - Revisited

Ali Rahimi and Ben Recht. Random Features for Large-Scale Kernel Machines. NIPS-2007

Zoltán Szabó

Machine Learning Journal Club, Gatsby

March 12, 2015

Notations

- $\mathbf{u}^T \in \mathbb{R}^d$: transpose.
- $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{R}^d} = \mathbf{v}^T \mathbf{u}$: inner product ($\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$).
- $\mathbf{u}^* \in \mathbb{C}^d$: conjugate transpose.
- $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathbb{C}^d} = \mathbf{v}^* \mathbf{u}$: inner product ($\mathbf{u}, \mathbf{v} \in \mathbb{C}^d$).
- $\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F$: spectral-, Frobenius norm.
- tr: trace.
- $Jf(\mathbf{a})$: Jacobi of function f at point $\mathbf{a}$.
- $\mathbb{E}$, $\mathbb{P}$: expectation, probability.
- $diam(S) = \sup\{\|\mathbf{a} - \mathbf{b}\|_2 : \mathbf{a}, \mathbf{b} \in S\}$: diameter.
  - Note: $diam(S) < \infty$ if $S \subseteq \mathbb{R}^d$, compact.
- $conv(S)$: convex hull of set $S \subseteq \mathbb{R}^d$.

Estimator

- Given: k shift-invariant kernel on $\mathbb{R}^d$, $k(\mathbf{x},\mathbf{y}) = \tilde{k}(\mathbf{x}-\mathbf{y})$.
- Bochner theorem: k cont. $\Rightarrow \exists p(\mathbf{w})$ density (k: normalized)

$$k(\mathbf{x},\mathbf{y}) = \int_{\mathbb{R}^d} p(\mathbf{w}) e^{i\mathbf{w}^T(\mathbf{x}-\mathbf{y})} \mathrm{d}\mathbf{w} = \mathbb{E}_{\mathbf{w}}[z_{\mathbf{w}}(\mathbf{x}) z_{\mathbf{w}}(\mathbf{y})^*], \quad z_{\mathbf{w}}(\mathbf{x}) = e^{i\mathbf{w}^T\mathbf{x}}.$$

- Estimator: $\mathbf{w}_1,\ldots,\mathbf{w}_D \overset{i.i.d.}{\sim} p(\mathbf{w})$,

$$\mathbf{z}(\mathbf{x}) = \frac{1}{\sqrt{D}}[z_{\mathbf{w}_1}(\mathbf{x});\ldots;z_{\mathbf{w}_D}(\mathbf{x})] \in \mathbb{C}^D,$$

$$s(\mathbf{x},\mathbf{y}) = \langle \mathbf{z}(\mathbf{x}),\mathbf{z}(\mathbf{y})\rangle_{\mathbb{C}^D} = \frac{1}{D}\sum_{j=1}^{D} e^{i\mathbf{w}_j^T(\mathbf{x}-\mathbf{y})} =: \tilde{s}(\mathbf{x}-\mathbf{y}) \in \mathbb{C},$$

$$\mathbb{E}[s(\mathbf{x},\mathbf{y})] = k(\mathbf{x},\mathbf{y}).$$

Cos features (issue-1)

In this case

$$z_{\mathbf{w},b}(\mathbf{x}) = \sqrt{2}\cos(\mathbf{w}^T\mathbf{x} + b) \in \mathbb{R}, \tag{1}$$

$$\begin{aligned} \mathbf{s}(\mathbf{x},\mathbf{y}) &= \langle \mathbf{z}(\mathbf{x}), \mathbf{z}(\mathbf{y}) \rangle_{\mathbb{R}^D} = \frac{2}{D}\sum_{j=1}^{D} \cos(\mathbf{w}_j^T\mathbf{x} + b_j)\cos(\mathbf{w}_j^T\mathbf{y} + b_j) \\ &= \frac{4}{D}\sum_{j=1}^{D} \cos(\mathbf{w}_j^T(\mathbf{x} - \mathbf{y})) + \cos(\mathbf{w}_j^T(\mathbf{x} + \mathbf{y}) + 2b_j) \end{aligned} \tag{2}$$

using $2\cos(a)\cos(b) = \cos(a-b) + \cos(a+b)$. Thus,

- $s(\mathbf{x},\mathbf{y})$: *not* translation invariant (assumed in the proof of Claim 1).
- We will also need vector-Hoeffding inequality $[s(\mathbf{x},\mathbf{y}) \in \mathbb{C}]$.

Goal, high-level proof

- **Goal** (uniform large deviation inequality):

$$\mathbb{P}\left(\sup_{\mathbf{x},\mathbf{y}\in\mathcal{M}} |s(\mathbf{x},\mathbf{y}) - k(\mathbf{x},\mathbf{y})| \geq \epsilon\right) \leq g(D,d,\epsilon). \qquad (3)$$

- Proof idea:
  - For fixed $(\mathbf{x},\mathbf{y})$:

$$s(\mathbf{x},\mathbf{y}) - k(\mathbf{x},\mathbf{y}) = \frac{1}{D}\sum_{j=1}^{D}\left[e^{i\mathbf{w}_j^T(\mathbf{x}-\mathbf{y})} - k(\mathbf{x},\mathbf{y})\right]. \qquad (4)$$

    $e^{i\mathbf{w}_j^T(\mathbf{x}-\mathbf{y})} - k(\mathbf{x},\mathbf{y})$: bounded $\Rightarrow$ concentration around 0 with high prob. (vector-Hoeffding).
  - 'ϵ-net' argument: extension from the net anchors (smoothness with high prob.).

Vector-Hoeffding inequality ⇒ step-1

Let $\{\xi_j\}_{j=1}^D \in H$(ilbert) i.i.d. and bounded ($\|\xi_j\|_H \le c$). Then

$$\mathbb{P}\left(\left\|\frac{1}{D}\sum_{j=1}^{D}\xi_j\right\|_H \ge \epsilon\right) \le 2e^{-\frac{D\epsilon^2}{2c^2}}. \tag{5}$$

In our case: $H = \mathbb{R}^2 (\cong \mathbb{C})$ and $c = 2$,

$$\begin{aligned}
\xi_j &= e^{i\mathbf{w}_j^T(\mathbf{x}-\mathbf{y})} - k(\mathbf{x},\mathbf{y}),\\
|\xi_j| &\le \left|e^{i\mathbf{w}_j^T(\mathbf{x}-\mathbf{y})}\right| + |k(\mathbf{x},\mathbf{y})| = 1 + |k(\mathbf{x},\mathbf{y})|,\\
|k(\mathbf{x},\mathbf{y})| &= \left|\int_{\mathbb{R}^d} p(\mathbf{w})e^{i\mathbf{w}^T(\mathbf{x}-\mathbf{y})}\mathrm{d}\mathbf{w}\right| \le \int_{\mathbb{R}^d}\left|p(\mathbf{w})e^{i\mathbf{w}^T(\mathbf{x}-\mathbf{y})}\right|\mathrm{d}\mathbf{w}\\
&= \int_{\mathbb{R}^d} p(\mathbf{w})\mathrm{d}\mathbf{w} = 1.
\end{aligned}$$

Step-2: net argument

Let $\mathbf{x}, \mathbf{y} \in \mathcal{M}$, $\mathcal{M}_{\mathbf{\Delta}} := \mathcal{M} - \mathcal{M} = \{\mathbf{x} - \mathbf{y} : \mathbf{x}, \mathbf{y} \in \mathcal{M}\} \subseteq \mathbb{R}^D$,

- $f(\mathbf{x}, \mathbf{y}) := s(\mathbf{x}, \mathbf{y}) - k(\mathbf{x}, \mathbf{y}) = \tilde{s}(\mathbf{x} - \mathbf{y}) - \tilde{k}(\mathbf{x} - \mathbf{y}) =: \tilde{f}(\mathbf{x} - \mathbf{y})$ $=: \tilde{f}(\mathbf{\Delta}) \in \mathbb{C}$, $\mathbf{\Delta} \in \mathcal{M}_{\mathbf{\Delta}}$.
- Assumption-1: $\mathcal{M}$ is compact. $\Rightarrow$
  - $\mathcal{M}_{\mathbf{\Delta}}$: compact,
  - $diam(\mathcal{M}_{\mathbf{\Delta}}) \leq 2diam(\mathcal{M})$,
  - $\mathcal{M}_{\mathbf{\Delta}}$: $\exists$ r-net with at most $N = \left(\frac{4diam(\mathcal{M})}{r}\right)^d$ balls of radius r,

$$\mathcal{M}_{\mathbf{\Delta}} \subseteq \cup_{\mathbf{\Delta}_j} B(\mathbf{\Delta}_j, r) = \cup_{\mathbf{\Delta}_j} \{\mathbf{u} : \|\mathbf{u} - \mathbf{\Delta}_j\|_2 < r\} \quad (j = 1, \ldots, N).$$

Step-2 (issue-2)

- Assumption-2: $\phi \circ \tilde{f} \in C^1$ $[\phi(v) = [\Re(v); \Im(v)]$, $v \in \mathbb{C}]$.
- Mean value theorem ($\tilde{f}(\mathbf{u}) \in \mathbb{C}$):

$$\phi\left[\tilde{f}(\mathbf{b}) - \tilde{f}(\mathbf{a})\right] = J\tilde{f}(\mathbf{c})(\mathbf{b} - \mathbf{a}), \quad [J\tilde{f}(\mathbf{c}) \in \mathbb{R}^{2\times D}] \tag{6}$$

  where $\mathbf{c} \in (\mathbf{a}, \mathbf{b}) = \{t\mathbf{a} + (1-t)\mathbf{b} : t \in (0,1)\}$. But this does *not* hold!
- Instead we have an integral variant ($\int$: meant coordinate-wise):

$$\phi\left[\tilde{f}(\mathbf{b}) - \tilde{f}(\mathbf{a})\right] = \left[\int_0^1 J(\phi \circ \tilde{f})(t\mathbf{a} + (1-t)\mathbf{b})\mathrm{d}t\right](\mathbf{b} - \mathbf{a}) \tag{7}$$

  assuming that $(\mathbf{a}, \mathbf{b}) \in \mathcal{M}_{\mathbf{\Delta}}$. $\Rightarrow$
- Assumption-3: $\mathbf{a}, \mathbf{b} \in \mathcal{M}_{\mathbf{\Delta}} \Rightarrow (\mathbf{a}, \mathbf{b}) \in \mathcal{M}_{\mathbf{\Delta}}$.

Step-2

$$
\begin{aligned}
|\tilde{f}(\mathbf{b}) - \tilde{f}(\mathbf{a})| &= \left\| \phi\left[\tilde{f}(\mathbf{b}) - \tilde{f}(\mathbf{a})\right]\right\|_2 \\
&\le \left\| \int_0^1 J(\phi\circ\tilde{f})(t\mathbf{a} + (1-t)\mathbf{b})\mathrm{d}t \right\|_2 \|\mathbf{b}-\mathbf{a}\|_2 \\
&\le \left(\int_0^1 \left\| J(\phi\circ\tilde{f})(t\mathbf{a} + (1-t)\mathbf{b})\right\|_2 \mathrm{d}t \right) \|\mathbf{b}-\mathbf{a}\|_2 \\
&\le \left(\int_0^1 L_{\tilde{f}} \mathrm{d}t \right) \|\mathbf{b}-\mathbf{a}\|_2 = L_{\tilde{f}} \|\mathbf{b}-\mathbf{a}\|_2,
\end{aligned}
$$

where

$$
L_{\tilde{f}} := \left\| J(\phi\circ\tilde{f})(\mathbf{\Delta}_*)\right\|_2, \mathbf{\Delta}_* := \arg\max_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \left\| J(\phi\circ\tilde{f})(\mathbf{\Delta})\right\|_2.
$$

Note: $\mathbf{\Delta}_*$ exists since $\mathcal{M}_{\mathbf{\Delta}}$ is compact, and $\phi\circ\tilde{f}\in C^1 \Rightarrow$ $\mathbf{\Delta}\mapsto \left\| J(\phi\circ\tilde{f})(\mathbf{\Delta})\right\|_2$ is continuous.

Step-2

- Lemma: if $|\tilde{f}(\mathbf{\Delta}_j)| < \frac{\epsilon}{2}$ $(\forall j)$, $L_{\tilde{f}} < \frac{\epsilon}{2r}$, then

$$|\tilde{f}(\mathbf{\Delta})| < \epsilon \quad (\forall \mathbf{\Delta} \in \mathcal{M}_{\mathbf{\Delta}}). \tag{8}$$

- Proof:

$$\left| \left|\tilde{f}(\mathbf{\Delta})\right| - \underbrace{\left|\tilde{f}(\mathbf{\Delta}_j)\right|}_{<\frac{\epsilon}{2}} \right| \le |\tilde{f}(\mathbf{\Delta}) - \tilde{f}(\mathbf{\Delta}_j)| \le \underbrace{L_{\tilde{f}}}_{\frac{\epsilon}{2r}} \underbrace{\|\mathbf{\Delta} - \mathbf{\Delta}_j\|_2}_{r} < \frac{\epsilon}{2}.$$

Thus, we got $|\tilde{f}(\mathbf{\Delta})| < \epsilon$.

Goal: guarantee the conditions of (8) with high probability.

Step-2a: guarantee $|\tilde{f}(\mathbf{\Delta}_j)| < \frac{\epsilon}{2}$ $(\forall j)$

Vector-Hoeffding inequality $\Rightarrow$

$$\mathbb{P}\left(|\tilde{f}(\mathbf{\Delta}_j)| \geq \frac{\epsilon}{2}\right) \leq 2e^{-\frac{D\epsilon^2}{2^2 8}} \Leftrightarrow \mathbb{P}\left(|\tilde{f}(\mathbf{\Delta}_j)| < \frac{\epsilon}{2}\right) \geq 1 - 2e^{-\frac{D\epsilon^2}{2^2 8}}.$$

By union bounding $(j = 1, \ldots, N)$:

$$\mathbb{P}\left(\cap_{j=1}^{N}\left\{|\tilde{f}(\mathbf{\Delta}_j)| < \frac{\epsilon}{2}\right\}\right) \geq 1 - 2Ne^{-\frac{D\epsilon^2}{32}}.$$

Step-2b: guarantee $L_{\tilde{f}} < \frac{\epsilon}{2r}$ (issue-3: $L_{\tilde{f}}$: $\mathbf{w}_j$-dep., $\|\cdot\|_2$)

- Markov inequality: $Z \ge 0$, $\mathbb{P}(Z \ge t) \le \frac{\mathbb{E}(Z)}{t}$ $(\forall t > 0)$. $Z = L_{\tilde{f}}$, $t = \frac{\epsilon}{2r}$.
- Bound on the expectation [$\|\mathbf{A}\|_2 \le \|\mathbf{A}\|_F$, $\|\sum_{j=1}^M f_j\|^2 \le M \sum_{j=1}^M \|f_j\|^2$, $\sup_{\mathbf{\Delta}}[f_1(\mathbf{\Delta}) + f_2(\mathbf{\Delta})] \le \sup_{\mathbf{\Delta}} f_1(\mathbf{\Delta}) + \sup_{\mathbf{\Delta}} f_2(\mathbf{\Delta})$]:

$$\mathbb{E}\left[L_{\tilde{f}}^2\right] = \mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \left\|J\left(\phi\circ\tilde{f}\right)(\mathbf{\Delta})\right\|_2^2\right] \le \mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \left\|J\left(\phi\circ\tilde{f}\right)(\mathbf{\Delta})\right\|_F^2\right]$$

$$= \mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \left\|J(\phi\circ\tilde{s})(\mathbf{\Delta}) - J\left(\phi\circ\tilde{k}\right)(\mathbf{\Delta})\right\|_F^2\right]$$

$$\le \mathbb{E}\left[2\sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \left(\|J(\phi\circ\tilde{s})(\mathbf{\Delta})\|_F^2 + \left\|J\left(\phi\circ\tilde{k}\right)(\mathbf{\Delta})\right\|_F^2\right)\right]$$

$$\le 2\mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \|J(\phi\circ\tilde{s})(\mathbf{\Delta})\|_F^2 + \sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \left\|J\left(\phi\circ\tilde{k}\right)(\mathbf{\Delta})\right\|_F^2\right]$$

$$=: 2\mathbb{E}\left[\sup_{\mathbf{\Delta}\in\mathcal{M}_{\mathbf{\Delta}}} \|J(\phi\circ\tilde{s})(\mathbf{\Delta})\|_F^2\right] + 2C_k,$$

Step-2b: continued ($\sigma_p^2 := \mathbb{E}_{\mathbf{w}\sim p}[\|\mathbf{w}\|_2^2]$)

$$\mathbb{E}(\|J(\phi\circ\tilde{s})(\boldsymbol{\Delta})\|_F^2) = \mathbb{E}\left(\left\|J\left[\phi\left(\frac{1}{D}\sum_{j=1}^{D}e^{i\mathbf{w}_j^T\boldsymbol{\Delta}}\right)\right]\right\|_F^2\right)$$

$$= \mathbb{E}\left(\left\|\frac{1}{D}\sum_{j=1}^{D}J\left[\phi\left(e^{i\mathbf{w}_j^T\boldsymbol{\Delta}}\right)\right]\right\|_F^2\right) = \frac{1}{D^2}\mathbb{E}\left(\left\|\sum_{j=1}^{D}J\left[\phi\left(e^{i\mathbf{w}_j^T\boldsymbol{\Delta}}\right)\right]\right\|_F^2\right)$$

$$\leq \frac{D}{D^2}\sum_{j=1}^{D}\mathbb{E}\left(\left\|J\left[\phi\left(e^{i\mathbf{w}_j^T\boldsymbol{\Delta}}\right)\right]\right\|_F^2\right) = \frac{1}{D}\sum_{j=1}^{D}\mathbb{E}\left(\|\mathbf{w}_j\|_F^2\right) = \frac{D}{D}\sigma_p^2 = \sigma_p^2,$$

$$\left\|\sum_{j=1}^{D}f_j\right\|^2 \leq D\sum_{j=1}^{D}\|f_j\|^2; \quad \left\|J\left[\phi(e^{i\mathbf{w}^T\mathbf{z}})\right]\right\|_F = \|\mathbf{w}\|_2 \quad \text{(next page).}$$

Step-2b: continued

$$\left\| J\left[\phi(e^{i\mathbf{w}^T\mathbf{z}})\right]\right\|_F = \left\| J[\cos(\mathbf{w}^T\mathbf{z});\sin(\mathbf{w}^T\mathbf{z})]\right\|_F$$
$$= \left\| [-\sin(\mathbf{w}^T\mathbf{z});\cos(\mathbf{w}^T\mathbf{z})]\mathbf{w}^T\right\|_F$$
$$= \left\| [-\sin(\mathbf{w}^T\mathbf{z});\cos(\mathbf{w}^T\mathbf{z})]\right\|_2 \|\mathbf{w}\|_2 = 1\times\|\mathbf{w}\|_2 = \|\mathbf{w}\|_2,$$
$$\left\|\mathbf{a}\mathbf{b}^T\right\|_F = \sqrt{tr\left[(\mathbf{a}\mathbf{b}^T)^T(\mathbf{a}\mathbf{b}^T)\right]} = \sqrt{tr\left[\mathbf{b}\mathbf{a}^T\mathbf{a}\mathbf{b}^T\right]} = \sqrt{tr\left[\mathbf{b}^T\mathbf{b}\mathbf{a}^T\mathbf{a}\right]}$$
$$= \sqrt{\|\mathbf{b}\|_2^2\,\|\mathbf{a}\|_2^2} = \|\mathbf{a}\|_2\,\|\mathbf{b}\|_2.$$

Step-2b: continued

- Until now: $E\left[L_{\tilde{f}}^2\right] \le 2(\sigma_p^2 + C_k)$. Target quantity ($\sqrt{t} = \frac{\epsilon}{2r}$):

$$\mathbb{P}\left(L_{\tilde{f}} \ge \sqrt{t}\right) = \mathbb{P}\left(L_{\tilde{f}}^2 \ge t\right) \overset{\text{Markov}}{\le} \frac{E\left[L_{\tilde{f}}^2\right]}{t} \le \frac{2(\sigma_p^2 + C_k)}{t},$$

$$\mathbb{P}\left(L_{\tilde{f}} \ge \frac{\epsilon}{2r}\right) \le \left(\frac{2r}{\epsilon}\right)^2 2(\sigma_p^2 + C_k) \Leftrightarrow \mathbb{P}\left(L_{\tilde{f}} < \frac{\epsilon}{2r}\right) \ge 1 - r.h.s.$$

- We proved earlier ($N = \left(\frac{4diam(\mathcal{M})}{r}\right)^d$):

$$\mathbb{P}\left(\cap_{j=1}^N \left\{|\tilde{f}(\boldsymbol{\Delta}_j)| < \frac{\epsilon}{2}\right\}\right) \ge 1 - 2Ne^{-\frac{D\epsilon^2}{32}}.$$

- By union bounding:

$$\mathbb{P}\left(\left\{L_{\tilde{f}} < \frac{\epsilon}{2r}\right\} \bigcap \cap_{j=1}^N \left\{|\tilde{f}(\boldsymbol{\Delta}_j)| < \frac{\epsilon}{2}\right\}\right) \ge$$

$$\ge 1 - 2\left(\frac{4diam(\mathcal{M})}{r}\right)^d e^{-\frac{D\epsilon^2}{32}} - \left(\frac{2r}{\epsilon}\right)^2 2(\sigma_p^2 + C_k).$$

Performance guarantee

$$\mathbb{P}\left(\sup_{\mathbf{x},\mathbf{y}\in\mathcal{M}}|s(\mathbf{x},\mathbf{y})-k(\mathbf{x},\mathbf{y})|<\epsilon\right)\geq 1-2\left(\frac{4diam(\mathcal{M})}{r}\right)^{d}e^{-\frac{D\epsilon^2}{32}}$$

$$-\left(\frac{2r}{\epsilon}\right)^{2}2(\sigma_p^2+C_k)$$

$$=1-\kappa_1 r^{-d}-\kappa_2 r^2 =: (*),$$

$$\kappa_1 = 2\left[4diam(\mathcal{M})\right]^d e^{-\frac{D\epsilon^2}{32}}, \quad \kappa_2=\left(\frac{2}{\epsilon}\right)^2 2(\sigma_p^2+C_k),$$

$$\kappa_1 r^{-d}=\kappa_2 r^2 \Leftrightarrow \frac{\kappa_1}{\kappa_2}=r^{d+2}\Leftrightarrow r=\left(\frac{\kappa_1}{\kappa_2}\right)^{\frac{1}{d+2}},$$

$$(*)=1-2\kappa_2\left(\frac{\kappa_1}{\kappa_2}\right)^{\frac{2}{d+2}}=1-2\kappa_1^{\frac{2}{d+2}}\kappa_2^{1-\frac{2}{d+2}=\frac{d+2-2}{d+2}=\frac{d}{d+2}}=$$

Performance guarantee

$$= 1 - 2\left(2\left[4 diam(\mathcal{M})\right]^d e^{-\frac{D\epsilon^2}{32}}\right)^{\frac{2}{d+2}} \left[\left(\frac{2}{\epsilon}\right)^2 2(\sigma_p^2 + C_k)\right]^{\frac{d}{d+2}}$$

$$= 1 - 2^{1+\frac{2}{d+2}+\frac{4d}{d+2}+\frac{3d}{d+2}} \left[\frac{\sqrt{\sigma_p^2 + C_k} diam(\mathcal{M})}{\epsilon}\right]^{\frac{2d}{d+2}} e^{-\frac{D\epsilon^2}{16(d+2)}}$$

$$\geq 1 - 2^8 \left[\frac{\sqrt{\sigma_p^2 + C_k} diam(\mathcal{M})}{\epsilon}\right]^2 e^{-\frac{D\epsilon^2}{16(d+2)}}$$

assuming $\frac{\sqrt{\sigma_p^2+C_k} diam(\mathcal{M})}{\epsilon} \geq 1$ ($\leftrightarrow a^x, a \geq 1$) and using

$h_1(d) = 1 + \frac{2}{d+2} + \frac{4d}{d+2} + \frac{3d}{d+2} = 8 - \frac{12}{d+2}$: increasing $\Rightarrow d = \infty$,

$h_2(d) = \frac{d}{d+2} = 1 - \frac{2}{d+2}$: increasing $\Rightarrow d = \infty$.

Fixed r.h.s.: solve for D

$$q = 2^8 \left[\frac{\sqrt{\sigma_p^2 + C_k diam(\mathcal{M})}}{\epsilon}\right]^2 e^{-\frac{D\epsilon^2}{16(d+2)}},$$

$$-\frac{D\epsilon^2}{16(d+2)} = \log\left(\frac{q}{2^8}\left[\frac{\sqrt{\sigma_p^2 + C_k diam(\mathcal{M})}}{\epsilon}\right]^{-2}\right),$$

$$D = -\frac{16(d+2)}{\epsilon^2}\log\left(\left[\left(\frac{q}{2^8}\right)^{-\frac{1}{2}}\right]^{-2}\left[\frac{\sqrt{\sigma_p^2 + C_k diam(\mathcal{M})}}{\epsilon}\right]^{-2}\right)$$

$$= \frac{32(d+2)}{\epsilon^2}\log\left[\sqrt{\frac{2^8}{q}}\frac{\sqrt{\sigma_p^2 + C_k diam(\mathcal{M})}}{\epsilon}\right].$$

We proved the theorem under the assumptions

- k conditions:
  - shift-invariance,
  - continuity, $\phi \circ \tilde{f} : \mathbb{R}^D \rightarrow \mathbb{R}^2$ is in C^1.
- $\mathcal{M}$ requirements:
  - $\mathcal{M}$: compact.
  - (*): $\mathbf{a}, \mathbf{b} \in \mathcal{M}_{\mathbf{\Delta}} = \mathcal{M} - \mathcal{M} \Rightarrow (\mathbf{a}, \mathbf{b}) \in \mathcal{M}_{\mathbf{\Delta}}$.

Note:

- $\phi \circ \tilde{f} = \phi \circ \tilde{s} + \phi \circ \tilde{k}$:
  - $\phi \circ \tilde{s}$: $\mathbf{z} \mapsto [\cos(\mathbf{w}^T\mathbf{z}); \sin(\mathbf{w}^T\mathbf{z}] \in C^1$,
  - $\phi \circ \tilde{k} \in C^1 \Leftrightarrow \mathbf{z} \mapsto \tilde{k}(\mathbf{z}) \in C^1$.
- (*) holds if $\mathcal{M}$ is convex using $[S_1 = \mathcal{M},\ S_2 = -\mathcal{M}]$

$$conv(S_1 + S_2) = conv(S_1) + conv(S_2).$$

Thank you for the attention!