Supplementary Material: Online Kernel Learning with a Near Optimal Sparsity Bound

L un Z an ZHANGLIJ MSU EDU

J n n Y YIJINFEN MSU EDU

on J n RONGJIN CSE MSU EDU

D pa n o Co p n an ng n ng gan a n a an ng A

n L n LIN M 8 MAILS TSINGHUA EDU CN

D pa n o A o a on ng a n B ng C na

X ao H XIAOFEIHE CAD ZJU EDU CN

a a o a o o CAD C Co g o Co p n Z ang n ng o 3 7 C na

A. Proof of Theorem 2

p o a o o n on n o ppo o o a op a g o n on a o n a p $\mathcal{T}_1 = \{(\mathbf{x_i}, y_i)\}_{\mathbf{i}=1}^{\mathbf{n}}$ $\langle \kappa(\mathbf{x_i}, \cdot), \kappa(\mathbf{x_j}, \cdot)\rangle_{\mathcal{H}_\kappa} = \delta_{\mathbf{ij}}$ an $y_\mathbf{i} \in \{ \ , - \ \}$ o a on on on g d po no a n $\kappa(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^\mathbf{T}\mathbf{y})^\mathbf{d}$ an an an pa $\mathbb{R}^\mathbf{m}$ $m > n$ n $m > n$ an n a o o ono a o $\{\mathbf{x}_1, \ldots, \mathbf{x_n}\}$ n $\mathbb{R}^\mathbf{m}$ a $\mathbf{x_i^T x_j} =$ n $i \neq j$ an $\mathbf{x_i^T x_i} =$ a o a on on a o a p on $\kappa(\mathbf{x_i}, \mathbf{x_j}) = \delta_{\mathbf{ij}}$ o a an n n an n $\mathbf{x_i}$ an $\mathbf{x_j}$ a g no g a o a $\kappa(\mathbf{x_i}, \mathbf{x_j}) \approx \delta_{\mathbf{ij}}$

Ba on $\mathcal{T}_1$ on ano $\mathcal{T}_2$ $(\ , u) \in \mathcal{T}_2$ an n $j \in n]$ an a n on $\xi \in \mathcal{H}_\mathbf{K}$ a

$$\kappa(\ , \cdot) = \kappa(\mathbf{x_j}, \cdot) \quad \xi, \ u = y_\mathbf{j}, \text{ an } \langle \xi, \kappa(\mathbf{x_i}, \cdot)\rangle_{\mathcal{H}_\kappa} = \ , \forall i \in \ n]. \tag{3}$$

o a $(\ , u) \in \mathcal{T}_2$ a o pon ng $(\mathbf{x_j}, y_\mathbf{j}) \in \mathcal{T}_1$ a a on p n (3) o n o $\mathcal{T}_2$ an p o n a a a a o $\mathcal{T}_1$

on T n a n ng a p $(\ _1, u_1), \ldots, (\ _\mathbf{T}, u_\mathbf{T})$ o $\mathcal{T}_2$ a o a $(\mathbf{x}, y) \in \mathcal{T}_1$ a T/n a p on o a ng og o $\ell(y, z) =$ n(p$(-yz)$) a an a p o a o on on a o a

$$f_* = \frac{R}{\sqrt{n}} \sum_{\mathbf{i}=1}^{\mathbf{n}} y_\mathbf{i} \kappa(\mathbf{x_i}, \cdot)$$

n a o on T a n ng a p

$$f_* = \underset{\|\mathbf{f}\|_{\mathcal{H}_\kappa} \leq \mathbf{R}}{\text{a g n}} \sum_{\mathbf{i}=1}^{\mathbf{T}} \ell(u_\mathbf{i}, f(\ _\mathbf{i})) = \underset{\|\mathbf{f}\|_{\mathcal{H}_\kappa} \leq \mathbf{R}}{\text{a g n}} \sum_{\mathbf{i}=1}^{\mathbf{T}} \ell(u_\mathbf{i}, \langle f, \kappa(\ _\mathbf{i}, \cdot)\rangle_{\mathcal{H}_\kappa}) = \underset{\|\mathbf{f}\|_{\mathcal{H}_\kappa} \leq \mathbf{R}}{\text{a g n}} \sum_{\mathbf{j}=1}^{\mathbf{n}} \sum_{\mathbf{k}=1}^{\mathbf{T/n}} \ell(y_\mathbf{j}, \langle f, \kappa(\mathbf{x_j}, \cdot) \quad \xi_{\mathbf{jk}}\rangle_{\mathcal{H}_\kappa}),$$

an n a o g n

$$\epsilon = T \ \text{n}(\quad \text{p}(-R/\sqrt{n})).$$

C oo ng

$$R = \sqrt{n} \ \text{n} \frac{T}{n},$$

$$\epsilon \leq T \quad \mathrm{p}(-R/\sqrt{n}) = n.$$

op o on f_* n ppo o an o o $O(n)$

n p o an o A go o o no o g an n o ppo o o on o o

$$O(\epsilon \quad R^2) = O(n \ \mathrm{n}T]^2).$$

na on an go o p q n o n $f'_1, \ldots, f'_{\mathbf{T}}$ no o an $n-$ ppo o B o on on go T/n n ng p an o $\sum_{\mathbf{i}=1}^{\mathbf{T}} \ell(u_{\mathbf{i}}, f'_{\mathbf{i}}(\ _{\mathbf{i}}))$ *larger than*

$$\frac{T}{n} \ \mathrm{n} \ = \ \left(\frac{T}{n}\right).$$

o f_* $O(n)$ o g o go o g an (T/n) gn an o an $O(n \ \mathrm{n}T]^2)$ o g T

B. Proof of Lemma 1

B n n n q o ng (C B an go) o n on o an

or *(Bernstein's inequality for martingales). Let $X_1, \ldots, X_{\mathbf{n}}$ be a bounded martingale difference sequence with respect to the filtration $\mathcal{F} = (\mathcal{F}_{\mathbf{i}})_{1 \leq \mathbf{i} \leq \mathbf{n}}$ and wit* $X_{\mathbf{i}}$

C. Proof of Lemma 2

[illegible] n on o X_t [illegible] K [illegible] n App n [illegible] B No [illegible] A_T n [illegible] pp o n o [illegible] [illegible] n o [illegible] [illegible] nno [illegible] app [illegible] o 3 o [illegible] [illegible] ng [illegible] o p ng p o [illegible] n (B [illegible] [illegible]) [illegible]

$$
\begin{aligned}
& \left(\quad \geq \sqrt{GA_{T}\tau} \quad \frac{}{3}K\tau \right) \\
= & \left(\quad \geq \sqrt{GA_{T}\tau} \quad \frac{}{3}K\tau, A_{T} \leq G_1 T \right) \\
= & \left(\quad \geq \sqrt{GA_{T}\tau} \quad \frac{}{3}K\tau, \quad {}^{2}_{T} \leq GA_{T}, A_{T} \leq G_1 T \right) \\
\leq & \left(\quad \geq \sqrt{GA_{T}\tau} \quad \frac{}{3}K\tau, \quad {}^{2}_{T} \leq GA_{T}, A_{T} \leq \frac{}{T} \right) \\
& \sum_{i=1}^{m} \left(\quad \geq \sqrt{GA_{T}\tau} \quad \frac{}{3}K\tau, \quad {}^{2}_{T} \leq GA_{T}, \frac{i-1}{T} < A_{T} \leq \frac{i}{T} \right) \\
\leq & \left(A_{T} \leq \frac{}{T} \right) \sum_{i=1}^{m} \left(\quad \geq \sqrt{\frac{G \quad i}{T}\tau} \quad \frac{}{3}K\tau, \quad {}^{2}_{T} \leq \frac{G \quad i}{T} \right) \\
\leq & \left(A_{T} \leq \frac{}{T} \right) \quad me^{-\tau},
\end{aligned}
$$

$m = \lceil \mathrm{og}_2(G_1 T^2) \rceil$ [illegible] n [illegible] p o o B n n n q [illegible] o [illegible] ng [illegible] o p p oo [illegible] ng $\tau =$ n(m/δ) [illegible] ng [illegible] p on $A_T >$ $/T$