A Detailed Information for Numerical Results

We provide the details of deep k-NN and mean and standard deviation information of figures in this section.

A.1 Deep k-NN

We follow the implementation of Papernot and McDaniel (2018) for the structures of the two neural networks as in Table 1:

Dataset	Network structure
MNIST	conv2d(1,16,3), conv2d(16,32,3), fc(32*7*7,32), fc(32,10)
CIFAR-10	conv2d(3,16,3), conv2d(16,16,3), conv2d(16,32,3), conv2d(32,32,3), fc(32*8*8,128), fc(128,10)

Table 1: CNN structure Summary

For both MNIST and CIFAR-10, we train the neural network using Adam optimizer with default learning rate, and use the output of all but the last fully-connected layer as features for k-NN. The number of neighbors is set to be 6. The number of epochs in training is 1 for MNIST and 50 for CIFAR-10.

A.2 Section 4.1.1 Figure 1

See Table 2.

n	Data	ω =0.02	ω =0.05	ω =0.1	ω =0.2	ω =0.5
64	clean	0.0686(0.0181)	0.0688(0.0181)	0.0693(0.0179)	0.0707(0.0177)	0.0787(0.0161)
	perturbed	0.069(0.0196)	0.072(0.0198)	0.0699(0.0195)	0.0735(0.0188)	0.0875(0.0153)
128	clean	0.0519(0.0198)	0.052(0.0198)	0.0525(0.0199)	0.0546(0.0197)	0.0668(0.0182)
	perturbed	0.0513(0.0166)	0.0543(0.0147)	0.052(0.0168)	0.0562(0.0136)	0.0735(0.0168)
256	clean	0.0388(0.0112)	0.0389(0.011)	0.0395(0.0108)	0.0422(0.0105)	0.0564(0.0094)
	perturbed	0.0384(0.0106)	0.0377(0.0105)	0.0391(0.0092)	0.0429(0.0087)	0.0619(0.0106)
512	clean	0.0274(0.0076)	0.0276(0.0075)	0.0285(0.0075)	0.0309(0.0072)	0.0478(0.0064)
	perturbed	0.0276(0.0083)	0.0261(0.0064)	0.0303(0.0094)	0.0325(0.0064)	0.0533(0.0069)
1024	clean	0.0201(0.006)	0.0204(0.0062)	0.0214(0.0061)	0.0245(0.0059)	0.0438(0.0055)
	perturbed	0.0199(0.0062)	0.0211(0.0066)	0.0225(0.0075)	0.0266(0.0067)	0.0496(0.007)
2048	clean	0.0139(0.0038)	0.014(0.0036)	0.0149(0.0037)	0.0187(0.0036)	0.0399(0.0038)
	perturbed	0.014(0.0036)	0.014(0.0034)	0.0155(0.0041)	0.0205(0.0044)	0.0428(0.0049)

Table 2: Mean and Standard Error of Regret for Figure 1

A.3 Section 4.1.1 Figure 2

See Table 3.

d	$\omega/\sqrt{d}$ =0.02	$\omega/\sqrt{d}$ =0.05	$\omega/\sqrt{d}$ =0.1	$\omega/\sqrt{d}$ =0.2	$\omega/\sqrt{d}$ =0.5
5	0.052(0.0199)	0.0525(0.0198)	0.0554(0.0197)	0.0642(0.0184)	0.1004(0.0176)
10	0.0796(0.02)	0.0797(0.0198)	0.0807(0.02)	0.0847(0.0198)	0.1035(0.0226)
15	0.0835(0.0125)	0.0836(0.0125)	0.0842(0.0121)	0.0868(0.0116)	0.0992(0.0118)
20	0.0929(0.0148)	0.093(0.0146)	0.0934(0.0143)	0.0947(0.0141)	0.1032(0.0142)
50	0.1038(0.0105)	0.1039(0.0104)	0.1038(0.0102)	0.1045(0.0107)	0.1087(0.0131)
100	0.1095(0.0143)	0.1096(0.0149)	0.1098(0.0152)	0.1101(0.0148)	0.113(0.0154)

Table 3: Mean and Standard Error of Regret for Figure 2

A.4 Section 4.1.2 Figure 3

Abalone: See Table 4.

	clean		perturbed	
ω	mean	std	mean	std
0	0.2192	0.0101	0.2192	0.0101
0.5	0.2322	0.0122	0.2373	0.0128
1	0.2567	0.0136	0.2598	0.0127
1.5	0.2761	0.0149	0.2726	0.0129
2	0.2837	0.0139	0.2801	0.0139
2.5	0.2980	0.0130	0.2877	0.0118
3	0.3042	0.0150	0.2932	0.0140

Table 4: Mean and Standard Error of Regret for Abalone dataset in Figure 3

HTRU2: See Table 5

	clean		perturbed	
omega	mean	std	mean	std
0	0.0225	0.0019	0.0229	0.0022
0.5	0.0263	0.0022	0.0259	0.0022
1	0.0337	0.0033	0.0288	0.0025
1.5	0.0439	0.0049	0.0316	0.0024
2	0.0561	0.0067	0.0357	0.0031
2.5	0.0697	0.0072	0.0390	0.0030
3	0.0835	0.0087	0.0419	0.0030

Table 5: Mean and Standard Error of Regret for HTRU2 dataset in Figure 3

A.5 Section 4.3.1 Figure 4

See Table 6.

n	Method	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 10$
64	pre-1nn	0.0485(0.0294)	0.0524(0.0392)	0.0747(0.0293)	0.0922(0.0329)	0.2199(0.0396)
	knn	0.0484(0.0275)	0.0477(0.0427)	0.0558(0.0342)	0.065(0.0361)	0.1531(0.0433)
128	pre-1nn	0.0351(0.029)	0.0431(0.0319)	0.0534(0.0282)	0.0722(0.026)	0.1859(0.0239)
	knn	0.0367(0.0296)	0.0417(0.0326)	0.0378(0.0252)	0.0409(0.016)	0.0902(0.0264)
256	pre-1nn	0.0275(0.026)	0.0222(0.0221)	0.027(0.0128)	0.0508(0.0237)	0.1615(0.0149)
	knn	0.0282(0.0258)	0.0174(0.0182)	0.0222(0.0178)	0.0277(0.0219)	0.0614(0.0161)
512	pre-1nn	0.0169(0.0179)	0.0124(0.0126)	0.0201(0.0134)	0.0306(0.0136)	0.1381(0.017)
	knn	0.013(0.0151)	0.0096(0.0145)	0.0114(0.0146)	0.014(0.0117)	0.0383(0.0147)
1024	pre-1nn	0.0077(0.0155)	0.0086(0.0118)	0.0125(0.0132)	0.0207(0.0114)	0.1208(0.0153)
	knn	0.0103(0.0175)	0.0076(0.0098)	0.0065(0.0105)	0.01(0.0092)	0.0223(0.0103)
2048	pre-1nn	0.0076(0.0108)	0.0037(0.0078)	0.0071(0.0098)	0.0123(0.0078)	0.1023(0.0127)
	knn	0.0073(0.0086)	0.0019(0.0076)	0.0055(0.0082)	0.0035(0.0066)	0.0106(0.0076)
4096	pre-1nn	0.0025(0.008)	0.0016(0.008)	0.005(0.0081)	0.0098(0.0087)	0.0864(0.013)
	knn	0.0018(0.0082)	0.0023(0.0073)	0.0024(0.0057)	0.0022(0.0053)	0.0063(0.0064)

Table 6: Mean and Standard Error of Regret for Figure 4

A.6 Section 4.3.2 Figure 5

See Table 7.

	knn		pre-1nn	
n	mean	std	mean	std
128	0.2781	0.0248	0.3160	0.0193
256	0.2079	0.0158	0.2455	0.0139
512	0.1584	0.0092	0.1852	0.0093
1024	0.1201	0.0052	0.1406	0.0051
2048	0.0915	0.0036	0.1068	0.0033
4096	0.0704	0.0021	0.0824	0.0025

Table 7: Mean and Standard Error of Regret for Figure 6

A.7 Section 4.3.2 Figure 6

See Table 8.

	knn		pre-1nn	
n	mean	std	mean	std
128	0.2650	0.0174	0.2771	0.0189
256	0.2478	0.0166	0.2588	0.0166
512	0.2360	0.0131	0.2465	0.0148
1024	0.2285	0.0130	0.2387	0.0116
2048	0.2212	0.0099	0.2306	0.0110

Table 8: Mean and Standard Error of Regret for Figure 7

A.8 Additional Real-Data Experiments

In Credit data set, there are 30000 samples (25% as testing data) with 23 attributes. For HTRU2 and Credit data set, the mean and standard error of error rates in the 50 repetitions are summarized in Table 9. From Table 9, using k-NN we obtain slightly smaller error rate on average.

Data Set	k-NN	Pre-1NN
Credit	0.1888(0.0041)	0.1900(0.0040)
HTRU2	0.0214(0.0021)	0.0221(0.0021)

Table 9: Mean and Standard Deviation of Error Rate using k-NN and Pro-processed 1NN (Pre-1NN) in HTRU2 and Credit. The error rate of pre-processed 1NN is always greater than that of k-NN.

B Comparison between Random Perturbation and Non-random Perturbation

We use the following adversarial attack as the non-random perturbation:

$$\widetilde{x} = \begin{cases} \underset{z\in B(x,\omega)}{\text{argmin}}\ \eta(z) & \text{if } \eta(x) > 1/2 \\ \underset{z\in B(x,\omega)}{\text{argmax}}\ \eta(z) & \text{if } \eta(x) \leq 1/2 \end{cases}. \tag{1}$$

When $\omega \to 0$, if η is differentiable, the length o attack converges to ω as well.

The proposed attack scheme (1) is also called as "white-box attack" as the adversary has the knowledge of $\eta(x)$. On the other hand, unlike the "white-box attack" mentioned in Wang et al. (2017), the perturbation and attack we focus on are independent with the training samples.

Theorem S. 1. *Under [A.1] to [A.3], if testing data is adversarially attacked and $1/\sqrt{k}, \zeta \ll \omega$, then*

$$Regret(k,n,\omega) = \frac{B_1}{4k} + \frac{1}{2}\int_{\mathcal{S}} \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2}\left(b(x_0)^2\zeta(x_0)^2 + 2\omega^2\|\dot{\eta}(x_0)\|^2\right) d\,Vol^{d-1}(x_0) + Rem, \tag{2}$$

where Rem $:= O(\omega/\sqrt{k} + \omega\zeta) + o((1/k) \vee (\zeta+\omega)^2)$.

From Theorem 1, one can see that the regret under adversarial attack is larger than the one under random perturbation if the ω^2 term is dominant.

C Relaxing Noise Distribution in Theorem 1

In Theorem 1, we assume the noise uniformly distributed on the sphere of a $\mathcal{L}_2$ with radius ω. However, as will be shown in the next section, we only utilize the distribution information of the noise at the last step of proof. F or general distribution of noise δ, we have the following result:

Theorem S. 2. *Under the same conditions as in Theorem 1. If* $\omega \to 0$ *in* n *such that* $P(\|\delta\| > \omega) = o(\omega^3)$*, then Theorem 1 holds as well.*

We connect the two ω's in Theorem S.2 and Theorem 1 using an example. Denote ω_0 as radius of $\mathcal{L}_2$ ball in Theorem 1, and assume $\omega_0 n^\gamma \to 0$ for some constant $\gamma > 0$. Take $\delta \sim N(0, \omega_0\Sigma)$ and the largest eigenvalue of Σ is finite. Then for some $c > 0$, we have $P(\|\delta\| > c\sqrt{d\log n}\omega_0) = o((\sqrt{d\log n}\omega_0)^3)$, i.e. the ω in Theorem S.2 in this case becomes $c\sqrt{d\log n}\omega_0$.

D Proof of Regret Analysis in Section 2 and 3

Sketch of Proof of Theorem 1 A sketch of proof is presented below.

Denote δ as the random perturbation, i,e., $\delta = \widetilde{x} - x$. Denote $X_1(x)$ to $X_k(x)$ be the k unsorted neighbors of x in the training samples and $Y_i(x)$ be the Y value for the corresponding $X_i(x)$. Similarly define $R_i(x)$ be the distance from x to $X_i(x)$. When no confusion is caused, we drop the argument x and use X_i, Y_i and R_i for abbreviation. The idea of proof follows Samworth (2012), and there are total 4 steps in our proof:

Step 1: Given a fixed (unobserved) testing sample x and conditional on the perturbation random variable δ, we obtain the mean and variance of $\widehat{\eta}_{k,n}(x+\delta)$. In particular, for any x_0 satisfying $\eta(x_0) = 1/2$, let $x_0^t = x_0 + t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}$ and $\epsilon \to 0$ such that $(1/k, R_1^2(x+\delta), \omega)$ are in $O(\epsilon)$, denoting $R_1 = R_1(x+\delta)$, we have

$$\begin{aligned}
\mathbb{E}[\widehat{\eta}_{k,n}(x_0^t+\delta)|\delta] &= \eta(x_0) + t\|\dot{\eta}(x_0)\| + \delta^\top\dot{\eta}(x_0) + b(x_0)R_1^2 + O(\epsilon^2),\\
Var(\widehat{\eta}_{k,n}(x_0^t+\delta)|\delta) &= \frac{1}{4k} + O(\epsilon^2/k).
\end{aligned}$$

Step 2: Use tube theory (Gray, 2012) to construct a tube for some $\mathcal{S}$. The remainder of regret outside the tube is of $O(\epsilon^3)$ for some ϵ, and Regret can be approximated as

$$\int_{\mathcal{S}}\int_{-\epsilon}^{\epsilon} t\|\dot{\Psi}(x_0)\|\mathbb{E}\big[P(\widehat{\eta}_{k,n}(x_0^t+\delta) < 1/2) - 1_{\{t<0\}}\big]dtd\text{Vol}^{d-1}(x_0).$$

Step 3: Use Berry-Esseen Theorem to transform the probability $P(\widehat{\eta}_{k,n}(x_0^t+\delta) < 1/2|\delta)$ to a Gaussian probability.

$$\Phi\left(\frac{\mathbb{E}(1/2 - \widehat{\eta}_{k,n}(x_0+\delta))}{\sqrt{Var(\widehat{\eta}_{k,n}(x_0^t+\delta))}}\bigg|\delta\right). \tag{3}$$

Step 4: Plug in the mean and variance of $\widehat{\eta}_{k,n}$ from *Step 1* into (3), integrate in the formula in *Step 2* on the tube over t. Finally, take expectation w.r.t. δ.

D.1 Theorem 1

This section contains the proof of Theorem 1 and Theorem S.1. The two proofs are similar, so for proof of Theorem S.1, we only present the part where the proof is different from Theorem 1.

Recall that $R_1(x)$ to $R_k(x)$ as the unsorted distance from the nearest k neighbors to testing data point x, and $R_{k+1}(x)$ as the distance from the exact $(k+1)$-th nearest neighbor to x itself. Similar as Chaudhuri and Dasgupta

(2014), conditional on the distance of the $(k+1)$-th neighbor, the first k neighbors are i.i.d. random variables distributed within $B(x, R_{k+1}(x))$.

In addition to f_1, f_2, and Ψ, we further denote $\bar{f}(x)$ as the density of X.

Proposition 3 (Lemma S.1 in Sun et al. (2016)). *For any distribution function G with density g,*

$$\begin{aligned}\int_{\mathbb{R}}[G(-bu-a)-1_{\{u<0\}}]du &= -\frac{1}{b}\left\{a+\int_{\mathbb{R}}tg(t)dt\right\},\\ \int_{\mathbb{R}}u[G(-bu-a)-1_{\{u<0\}}]du &= \frac{1}{b^2}\left\{\frac{a^2}{2}+\frac{1}{2}\int_{\mathbb{R}}t^2g(t)dt+a\int_{\mathbb{R}}tg(t)dt\right\}.\end{aligned}$$

Now we start our proof of Theorem 1.

D.1.1 Formal Proof

Proof of Theorem 1. Step 1: For the scenario of random perturbation, δ is a random variable uniformly distributed on sphere $B(x_0^t, \omega)$, we first evaluate $\mathbb{E}(\widehat{\eta}_{k,n}(x_0^t+\delta))$ and $Var(\widehat{\eta}_{k,n}(x_0^t+\delta))$ for given x_0 and δ.

$$\begin{aligned}&\mathbb{E}[\widehat{\eta}_{k,n}(x_0^t+\delta)|\delta] = \mathbb{E}(Y_1(x_0^t+\delta)|\delta)\\ =&\mathbb{E}\left(\eta(x_0^t+\delta)+(X_1-x_0^t-\delta)^\top\dot{\eta}(x_0^t+\delta)+1/2(X_1-x_0^t-\delta)^\top\ddot{\eta}(x_0^t+\delta)(X_1-x_0^t-\delta)\middle|\delta\right)+rem,\end{aligned}$$

where rem is a remainder term due to the Taylor's expansion. Given δ and $R_1(x_t^0+\delta)=\|X_1-x_0^t-\delta\|$, the distribution of X_1 is on the sphere of $B(x_t^0+\delta, R_1)$. Denote the density of this distribution as $\bar{f}(x|x_0^t+\delta, R_1(x_0^t+\delta))$. Also define $\bar{f}'(x|x_0^t+\delta, R_1(x_0^t+\delta))$ as the gradient of $\bar{f}(x|x_0^t+\delta, R_1(x_0^t+\delta))$. For simplicity, rewrite $R_1(x_0^t+\delta)$ as R_1. Then based on (A.1) and (A.3) for the smoothness of $\bar{f}$ and η, rewrite $\bar{f}(x|x_0^t+\delta, R_1)$ as a Taylor expansion at $x_0^t+\delta$, and we have

$$\begin{aligned}&\mathbb{E}((X_1-x_0^t-\delta)^\top\dot{\eta}(x_0^t+\delta)|\delta, R_1)\\ =&\int_{\partial B}(x-x_0^t-\delta)^\top\dot{\eta}(x_0^t+\delta)\bar{f}(x|x_0^t+\delta, R_1)dx\\ =&\int_{\partial B}(x-x_0^t-\delta)^\top\dot{\eta}(x_0^t+\delta)\Big[\bar{f}(x_0^t+\delta|x_0^t+\delta, R_1)\\ &\qquad+\bar{f}'(x_0^t+\delta|x_0^t+\delta, R_1)^\top(x-x_0^t-\delta)\\ &\qquad+\frac{1}{2}(x-x_0^t-\delta)^\top\bar{f}''(x_0^t+\delta|x_0^t+\delta, R_1)(x-x_0^t-\delta)\\ &\qquad+O(\|x-x_0^t-\delta\|_2^3)\Big]dx\\ =&\int_{\partial B}(x-x_0^t-\delta)^\top\dot{\eta}(x_0^t+\delta)\bar{f}'(x_0^t+\delta|x_0^t+\delta, R_1)^\top(x-x_0^t-\delta)dx+o\\ =&\ tr\left(\dot{\eta}(x_0^t+\delta)\bar{f}'(x_0^t+\delta|x_0^t+\delta, R_1)^\top\int_{\partial B}(x-x_0^t-\delta)(x-x_0^t-\delta)^\top dx\right)+O(R_1^4),\end{aligned}$$

where $\int_{\partial B}$ denotes integration over sphere $\partial B(x_0^t+\delta, R_1)$ the first-order and third-order terms becomes 0.

In addition,

$$\begin{aligned}&tr\left(\frac{1}{2}\ddot{\eta}(x_0^t+\delta)\mathbb{E}\left((X_1-x_0^t-\delta)(X_1-x_0^t-\delta)^\top|R_1\right)\right)\\ =&\ tr\left(\frac{1}{2}\ddot{\eta}(x_0^t+\delta)\int_{\partial B}(x-x_0^t-\delta)(x-x_0^t-\delta)^\top\bar{f}(x|x_0^t+\delta, R_1)dx\right)\\ =&\ tr\left(\frac{1}{2}\ddot{\eta}(x_0^t+\delta)\int_{\partial B}(x-x_0^t-\delta)(x-x_0^t-\delta)^\top\bar{f}(x_0^t+\delta|x_0^t+\delta, R_1)dx\right)\\ &+tr\left(\frac{1}{2}\ddot{\eta}(x_0^t+\delta)\int_{\partial B}(x-x_0^t-\delta)(x-x_0^t-\delta)^\top(x-x_0^t-\delta)^\top\bar{f}'(x_0^t+\delta|x_0^t+\delta, R_1)dx\right)\\ &+O(R_1^4).\end{aligned}$$

The term rem in $\mathbb{E}(\widehat{\eta}_{k,n})$ can be tackled in a similar manner and $rem = O(R_1^4)$. Hence taking

$$b(x) = \frac{1}{\bar{f}(x)d}\left\{\sum_{j=1}^{d}[\dot{\eta}_j(x)\dot{\bar{f}}_j(x) + \ddot{\eta}_{j,j}(x)\bar{f}(x)/2]\right\},$$

we have

$$\begin{aligned}
\mathbb{E}(\widehat{\eta}_{k,n}|\delta, R_1) &= \eta(x_0^t+\delta) + b(x_0^t+\delta)R_1^2 + O(R_1^4) \\
&= \eta(x_0) + \frac{t}{\|\dot{\eta}(x_0)\|}\dot{\eta}(x_0)^\top\dot{\eta}(x_0) + \delta^\top\dot{\eta}(x_0) + O(t^2+\omega^2) \\
&\quad +b(x_0)R_1^2 + R_1^2\frac{t}{\|\dot{\eta}(x_0)\|}\dot{\eta}(x_0)^\top\dot{b}(x_0) + R_1^2\delta^\top\dot{b}(x_0) + O(R_1^4) \\
&= \eta(x_0) + t\|\dot{\eta}(x_0)\| + \delta^\top\dot{\eta}(x_0) + b(x_0)R_1^2 + O(t^2+\omega^2+R_1^4).
\end{aligned}$$

Denote $t_{k,n}(x_0^t+\delta) = \mathbb{E}R_1^2$, using arguments in Lemma 1 and Theorem 2 of Xing et al. (2018), take $a_d = 2^d\Gamma(1+1/2)^d/\Gamma(1+d/2)$, we obtain

$$\begin{aligned}
t_{k,n}(x_0^t+\delta) &= \frac{1}{a_d^{2/d}\bar{f}(x_0^t+\delta)^{2/d}}\left(\frac{k}{n}\right)^{2/d} + o(t_{k,n}^2(x_0^t+\delta)) \\
&= t_{k,n}(x_0) + O\left(t\left(\frac{k}{n}\right)^{2/d}\right) + O\left(\omega\left(\frac{k}{n}\right)^{2/d}\right) + o(t_{k,n}^2(x_0^t+\delta)).
\end{aligned}$$

Further denote $\mu_{k,n,\omega}(x_0^t,\delta) = \eta(x_0) + t\|\dot{\eta}(x_0)\| + \delta^\top\dot{\eta}(x_0) + b(x_0)t_{k,n}(x_0)$, we obtain

$$\mathbb{E}(\widehat{\eta}_{k,n}|\delta) = \mu_{k,n,\omega}(x_0^t,\delta) + O(t^2+\omega^2+t_{k,n}^2) = \mu_{k,n,\omega}(x_0^t,\delta) + O(\epsilon_{k,n,\omega}^2).$$

In terms of $Var(\widehat{\eta}_{k,n}(x_0^t,\delta))$, fixing R_{k+1}, the k neighbors are i.i.d. random variables in $B(x_0^t+\delta, R_{k+1})$,

$$Var(Y_1|R_{k+1},\delta) = \mathbb{E}(Y_1|R_{k+1},\delta)(1-\mathbb{E}(Y_1|R_{k+1}\delta)) = \frac{1}{4} + O\left(\epsilon_{k,n,\omega}^2\right),$$

when $R_{k+1}^2 = O(t_{k,n}(x_0))$. Moreover, as Chaudhuri and Dasgupta (2014) and Belkin et al. (2018) mentioned, the probability of $R_{k+1} \gg t_{k,n}(x_0)$ is an exponential tail, hence the overall variance becomes

$$Var(Y_1|\delta) = \frac{1}{4} + O\left(\epsilon_{k,n,\omega}^2\right).$$

This also implies that

$$|\sqrt{Var(Y_1|\delta)} - \sqrt{1/4}| = O(\epsilon_{k,n,\omega}).$$

Step 2: Our aim is to quantify the following quantity:

$$\int_{\mathbb{R}^d}\Psi(x)\left(P\left(\sum_{i=1}^{k}\frac{1}{k}Y_i \le \frac{1}{2}\right) - 1_{\{\eta(x)<1/2\}}\right)dx,$$

which equals to

$$\int_{\mathcal{R}}\Psi(x)\left(P\left(\sum_{i=1}^{k}\frac{1}{k}Y_i \le \frac{1}{2}\right) - 1_{\{\eta(x)<1/2\}}\right)dx,$$

where $\mathcal{R}$ is the support of X. Taking $\epsilon_{k,n,\omega} \ge -s_{k,n}\log s_{k,n}$, we have

$$\begin{aligned}
&\int_{\mathbb{R}^d}\Psi(x)\left(P\left(\sum_{i=1}^{k}\frac{1}{k}Y_i \le \frac{1}{2}\middle|\delta\right) - 1_{\{\eta(x)<1/2\}}\right)dx \\
= &\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left[P(\widehat{\eta}_{k,n}(x_0^t+\delta) < 1/2) - 1_{\{t<0\}}\right]dtd\text{Vol}^{d-1}(x_0) + r_1. \qquad (4)
\end{aligned}$$

The result in (4) adopts tube theory Gray (2012) to transform the integration from $\mathbb{R}^d$ to $\mathbb{R} \times \mathcal{S}$. Denote the map $\phi\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right) = x_0^t$, then the pullback of the d-form dx is given at $(x_0, t\dot{\eta}(x_0)/\|\dot{\eta}(x_0)\|)$ by

$$det\left(\dot{\phi}\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right)\right) dtd\mathrm{Vol}^{d-1}(x_0).$$

For r_1, it is composed of four parts: (1) the integral outside $\mathcal{S}^{\epsilon_{k,n,\omega}}$, (2) the difference between $\Psi(t)$ and $t\|\dot{\Psi}(x)\|$, (3) the difference between $\mathcal{S}^{\epsilon_{k,n,\omega}}$ and the tube generated using $\mathcal{S}$, and (4) the remainder of $det\left(\dot{\phi}\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right)\right)$:

$$\begin{aligned}
&r_1 \\
=&\int_{\mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}}} \left(P\left(\sum_{i=1}^k \frac{1}{k} Y_i \leq \frac{1}{2}\middle| \delta\right) - 1_{\{\eta(x)<1/2\}}\right) dP(x) + O(\epsilon^3_{k,n,w}) \\
&+ \int_{\mathcal{S}} \int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\| \left(P(\widehat{\eta}_{k,n}(x_0^t + \delta) < 1/2) - 1_{\{t<0\}}\right) \left[det\left(\dot{\phi}\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right)\right) - 1\right] dtd\mathrm{Vol}^{d-1}(x_0) \\
:=& r_{11} + r_{12} + O(\epsilon^3_{k,n,w}).
\end{aligned} \tag{5}$$

For r_{11} in r_1:

$$\begin{aligned}
0 \;\; \geq& \;\; \int_{\mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}} \cap \{x|\eta(x)<1/2\}} \left(P\left(\sum_{i=1}^k \frac{1}{k} Y_i \leq \frac{1}{2}\middle| \delta\right) - 1_{\{\eta(x)<1/2\}}\right) dP(x) \\
=& \;\; \int_{\mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}} \cap \{x|\eta(x)<1/2\}} \left(P\left(\sum_{i=1}^k \frac{1}{k}\left(Y_i - \frac{1}{2}\right) - \mathbb{E}\left(Y_1 - \frac{1}{2}\right) \leq -\mathbb{E}\left(Y_1 - \frac{1}{2}\right)\middle| \delta\right) - 1_{\{\eta(x)<1/2\}}\right) dP(x) \\
=& \;\; -\int_{\mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}} \cap \{x|\eta(x)<1/2\}} P\left(\sum_{i=1}^k \frac{1}{k}\left(Y_i - \frac{1}{2}\right) - \mathbb{E}\left(Y_1 - \frac{1}{2}\right) > -\mathbb{E}\left(Y_1 - \frac{1}{2}\right)\middle| \delta\right) dP(x).
\end{aligned}$$

From the definition of $\epsilon_{k,n,\omega}$, we know that for any δ, $\inf\limits_{x \in \mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}}} |\mathbb{E}Y(\widetilde{x}_\omega) - 1/2| \geq c_1 \epsilon_{k,n,\omega}$ for some $c_1 > 0$. Using Berstein inequality, we have an upper bound as

$$\begin{aligned}
& \int_{\mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}} \cap \{x|\eta(x)<1/2\}} P\left(\sum_{i=1}^k \frac{1}{k}(Y_i - \frac{1}{2}) - \mathbb{E}(Y_1 - \frac{1}{2}) > -\mathbb{E}(Y_1 - \frac{1}{2})\middle| \delta\right) dP(x) \\
\leq & \;\; O(\exp(-c_2 k \epsilon^2_{k,n,\omega})) = o(1/k^{3/2}),
\end{aligned}$$

for $c_2 > 0$.

Similar result can be obtained for $\mathbb{R}^d \backslash \mathcal{S}^{\epsilon_{k,n,\omega}} \cap \{x|\eta(x) > 1/2\}$.

For r_{12} in r_1,

$$\begin{aligned}
& \int_{\mathcal{S}} \int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\| \left(P(\widehat{\eta}_{k,n}(x_0^t + \delta) < 1/2) - 1_{\{t<0\}}\right) \\
& \qquad \left[det\left(\dot{\phi}\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right)\right) - 1\right] dtd\mathrm{Vol}^{d-1}(x_0) \\
= & \int_{\mathcal{S}} \int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\| \left(P(\widehat{\eta}_{k,n}(x_0 + \delta) < 1/2) - 1_{\{t<0\}}\right) \\
& \qquad \left[det\left(\dot{\phi}\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right)\right) - 1\right] dtd\mathrm{Vol}^{d-1}(x_0) \\
& + \int_{\mathcal{S}} \int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\phi}(x_0)\| \left[t\frac{\dot{\eta}(x_0)^\top}{\|\dot{\eta}(x_0)\|} \frac{\partial}{\partial x_0}\left(P(\widehat{\eta}_{k,n}(x_0 + \delta) < 1/2) - 1_{\{t<0\}}\right)\right] \\
& \qquad \left[det\left(\dot{\phi}\left(x_0, t\frac{\dot{\eta}(x_0)}{\|\dot{\eta}(x_0)\|}\right)\right) - 1\right] dtd\mathrm{Vol}^{d-1}(x_0) + o \\
= & \;\; O(\epsilon^3_{k,n,\omega}).
\end{aligned}$$

Finally $r_1 = O(\epsilon^3_{k,n,\omega})$.

Step 3: we continue the derivation of

$$\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left(P(\widehat{\eta}_{k,n}(x_0^t+\delta)<1/2)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0).$$

Since given δ and R_{k+1}, $\widehat{\eta}_{k,n}$ is obtained from k i.i.d. samples (though for some samples their weight is 0), by non-uniform Berry-Esseen Theorem,

$$\begin{aligned}
&\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\mathbb{E}_{R_{k+1}}\left(P(\widehat{\eta}_{k,n}(x_0^t+\delta)<1/2)-1_{\{t<0\}}|\delta,R_{k+1}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
= &\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left(\Phi\left(\frac{k\mathbb{E}(1/2-Y_1)}{\sqrt{kVar(Y_1)}}\middle|\delta\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)+r_2\\
&+\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left(\mathbb{E}_{R_{k+1}}\Phi\left(\frac{k\mathbb{E}(1/2-Y_1|R_{k+1})}{\sqrt{kVar(Y_1|R_{k+1})}}\middle|\delta\right)-\Phi\left(\frac{k\mathbb{E}(1/2-Y_1)}{\sqrt{kVar(Y_1)}}\middle|\delta\right)\right)dtd\mathrm{Vol}^{d-1}(x_0).
\end{aligned}$$

where

$$\left|P(\widehat{\eta}_{k,n}(x_0^t+\delta)<1/2|R_{k+1},\delta)-\Phi\left(\frac{k\mathbb{E}(1/2-Y_1|R_{k+1})}{\sqrt{kVar(Y_1|R_{k+1})}}\middle|\delta\right)\right| \quad\leq\quad \frac{c_3}{\sqrt{k}}\frac{1}{1+k^{3/2}\left|\mathbb{E}1/2-Y_1|R_{k+1}\right|^3},$$

hence

$$r_2 \leq \int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\frac{c_3}{\sqrt{k}}\frac{1}{1+k^{3/2}\left|\mathbb{E}1/2-Y_1|R_{k+1}\right|^3}dtd\mathrm{Vol}^{d-1}(x_0)=O\left(\frac{1}{k^{3/2}}\right).$$

We can also obtain that

$$\begin{aligned}
&\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left(\mathbb{E}_{R_{k+1}}\Phi\left(\frac{k\mathbb{E}(1/2-Y_1|R_{k+1})}{\sqrt{kVar(Y_1|R_{k+1})}}\middle|\delta\right)-\Phi\left(\frac{k\mathbb{E}(1/2-Y_1)}{\sqrt{kVar(Y_1)}}\middle|\delta\right)\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
= &\ O(\epsilon^3_{k,n,\omega}).
\end{aligned}$$

Step 4: Finally we integrate on Gaussian probabilities:

$$\begin{aligned}
&\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left(\Phi\left(\frac{k\mathbb{E}(1/2-Y_1)}{\sqrt{kVar(Y_1)}}\middle|\delta\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
= &\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\left(\Phi\left(-\frac{t\|\dot{\eta}(x_0)\|}{\sqrt{s^2_{k,n}}}-\frac{(b(x_0)t_{k,n}(x_0)+\delta^\top\dot{\eta}(x_0^t))}{\sqrt{s^2_{k,n}}}\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
&+r_3\\
= &\int_{\mathcal{S}}\int_{\mathbb{R}} t\|\dot{\Psi}(x_0)\|\left(\Phi\left(-\frac{t\|\dot{\eta}(x_0)\|}{\sqrt{s^2_{k,n}}}-\frac{(b(x_0)t_{k,n}(x_0)+\delta^\top\dot{\eta}(x_0))}{\sqrt{s^2_{k,n}}}\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
&+r_3+r_4\\
= &\ \frac{1}{2}\int_{\mathcal{S}}\frac{1}{4k}\frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2}s^2_{k,n}d\mathrm{Vol}^{d-1}(x_0)+\int_S\frac{\|\dot{\Psi}(x_0)\|}{2\|\dot{\eta}(x_0)\|^2}(b(x_0)t_{k,n}(x_0)+\delta^\top\dot{\eta}(x_0))^2d\mathrm{Vol}^{d-1}(x_0)+r_3+r_4.
\end{aligned}$$

The last step follows Proposition 3. For the small order terms,

$$\begin{aligned}
r_3 \quad = &\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot{\Psi}(x_0)\|\Bigg(\Phi\left(\frac{k\mathbb{E}(1/2-Y_1)}{\sqrt{kVar(Y_1)}}\middle|\delta\right)\\
&-\Phi\left(-\frac{t\|\dot{\eta}(x_0)\|}{\sqrt{s^2_{k,n}}}-\frac{(b(x_0)t_{k,n}(x_0)+\delta^\top\dot{\eta}(x_0^t))}{\sqrt{s^2_{k,n}}}\right)\Bigg)dtd\mathrm{Vol}^{d-1}(x_0)\\
= &\ O(\epsilon^3_{k,n,\omega}).
\end{aligned}$$

Through definition of $\epsilon_{k,n,\omega}$ we have

$$r_4 \quad = \quad o(1/k^{3/2}).$$

Finally we take expectation on δ:

$$\begin{aligned}
& \mathbb{E}_\delta \int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2} (b(x_0)t_{k,n}(x_0) + \delta^\top \dot{\eta}(x_0))^2 d\mathrm{Vol}^{d-1}(x_0) \\
= & \int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2} \mathbb{E}_\delta (b(x_0)t_{k,n}(x_0) + \delta^\top \dot{\eta}(x_0))^2 d\mathrm{Vol}^{d-1}(x_0) \\
= & \int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2} \left(b(x_0)^2 t_{k,n}^2(x_0) + 2b(x_0)t_{k,n}(x_0)\mathbb{E}_\delta(\delta^\top \dot{\eta}(x_0)) + \mathbb{E}_\delta(\delta^\top \dot{\eta}(x_0))^2\right) d\mathrm{Vol}^{d-1}(x_0) \\
= & \int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2} \left(b^2(x_0)t_{k,n}^2(x_0) + \frac{\|\dot{\eta}(x_0)\|^2}{d}\omega^2\right) d\mathrm{Vol}^{d-1}(x_0).
\end{aligned}$$

□

D.2 Theorem 3

Proof of Theorem 3. When $|\eta(x) - 1/2| > C\omega$ for some large constant $C > 0$, g and $\widetilde{g}$ will always be the same, thus

$$\begin{aligned}
& P(\widetilde{g}(\widetilde{x}) \neq Y) - P(g(x) \neq Y) && (6) \\
= \quad & \mathbb{E}_\delta \left[\int_{\mathcal{S}} \int_{-C\omega}^{C\omega} t\|\dot{\Psi}(x_0)\| \left(1_{\{\widetilde{\eta}(x_0^t+\delta)<1/2\}} - 1_{\{t<0\}}\right) dt d\mathrm{Vol}^{d-1}(x_0) \right] + O(\omega^4). && (7)
\end{aligned}$$

Moreover,

$$\widetilde{\eta}(\widetilde{x}) = \mathbb{E}(\eta(x)|\widetilde{x} \text{ is observed}) = \eta(\widetilde{x}) + b(\widetilde{x})\omega^2 + O(\omega^3). \tag{8}$$

As a result,

$$\widetilde{\eta}(x_0^t + \delta) = \eta(x_0) + t\|\dot{\eta}(x_0)\| + \dot{\eta}(x_0)^\top \delta + b(x_0)\omega^2 + O(t\omega^2) + O(\omega^3). \tag{9}$$

Plugging in $\widetilde{\eta}(x_0^t + \delta)$ into regret, we obtain that

$$\begin{aligned}
& P(\widetilde{g}(\widetilde{x}) \neq Y) - P(g(x) \neq Y) && (10) \\
= \quad & \mathbb{E}_\delta \left[\int_{\mathcal{S}} \int_{-C\omega}^{C\omega} t\|\dot{\Psi}(x_0)\| \left(1_{\{\widetilde{\eta}(x_0^t+\delta)<1/2\}} - 1_{\{t<0\}}\right) dt d\mathrm{Vol}^{d-1}(x_0) \right] + O(\omega^4) && (11) \\
= \quad & \mathbb{E}_\delta \left[\int_{\mathcal{S}} \int_{-C\omega}^{C\omega} t\|\dot{\Psi}(x_0)\| \left(1_{\{t<-\dot{\eta}(x_0)^\top\delta/\|\dot{\eta}(x_0)\|-b(x_0)\omega^2\}} - 1_{\{t<0\}}\right) dt d\mathrm{Vol}^{d-1}(x_0) \right] + O(\omega^4) && (12) \\
= \quad & \mathbb{E} \left[\int_{\mathcal{S}} \|\dot{\Psi}(x_0)\| \int_{-\dot{\eta}(x_0)^\top\delta/\|\dot{\eta}(x_0)\|-b(x_0)\omega^2}^{0} t dt d\mathrm{Vol}^{d-1}(x_0) \right] + O(\omega^4) && (13) \\
= \quad & \int_{\mathcal{S}} \|\dot{\Psi}(x_0)\| \frac{\omega^2}{2d} d\mathrm{Vol}^{d-1}(x_0) + O(\omega^4). && (14)
\end{aligned}$$

From this derivation, the dominant terms in the denominator and numerator of the quantity

$$\frac{P(Y \neq \widehat{g}_n(\widetilde{x})) - P(Y \neq \widetilde{g}(\widetilde{x}))}{P(Y \neq \widehat{g}_n'(\widetilde{x})) - P(Y \neq \widetilde{g}(\widetilde{x}))} \tag{15}$$

are both $\Theta(n^{-4/(d+4)})$ when k's are chosen to be optimal respectively. Note that the multiplicative constants for numerator and denominator are both determined by δ and density of X, and converges to each other when $\omega \to 0$. As $\omega \to 0$ when $n \to \infty$, the difference on the densities vanishes, thus (15) converges to 1. □

D.3 Theorem S.1

Proof of Theorem S.1. The proof is similar with Theorem 1. Since the format of r_1 to r_4 are unchanged, one can show that they are small order terms in Theorem 3 as well. What is changed in the proof of Theorem 3 is $\mu_{k,n,\omega}(x)$:

When $t < 0$, we have

$$\mu_{k,n,\omega}(x_0^t) = \eta(x_0) + t\|\dot\eta(x_0)\| + \omega\|\dot\eta(x_0)\| + b(x_0)t_{k,n}(x) + o,$$

while for $t > 0$,

$$\mu_{k,n,\omega}(x_0^t) = \eta(x_0) + t\|\dot\eta(x_0)\| - \omega\|\dot\eta(x_0)\| + b(x_0)t_{k,n}(x) + o.$$

Therefore,

$$\begin{aligned}
&\int_{\mathcal{S}}\int_{-\epsilon_{k,n,\omega}}^{\epsilon_{k,n,\omega}} t\|\dot\Psi(x_0)\|\left(\Phi\left(\frac{k\mathbb{E}(1/2-Y_1)}{\sqrt{kVar(Y_1)}}\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
= &\int_{\mathcal{S}}\int_{\mathbb{R}} t\|\dot\Psi(x_0)\|\left(\Phi\left(-\frac{t\|\dot\eta(x_0)\|-sign(t)\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0))}{\sqrt{s_{k,n}^2}}\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)+o\\
= &\int_{\mathcal{S}}\int_{\mathbb{R}} t\|\dot\Psi(x_0)\|\left(\Phi\left(-\frac{t\|\dot\eta(x_0)\|+\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0))}{\sqrt{s_{k,n}^2}}\right)-1_{\{t<0\}}\right)dtd\mathrm{Vol}^{d-1}(x_0)+r_5+o\\
= &\frac{1}{2}\int_{\mathcal{S}}\frac{1}{4k}\frac{\|\dot\Psi(x_0)\|}{\|\dot\eta(x_0)\|^2}s_{k,n}^2 d\mathrm{Vol}^{d-1}(x_0)+\frac{1}{2}\int_S\frac{\|\dot\Psi(x_0)\|}{\|\dot\eta(x_0)\|^2}\left(b(x_0)\mathbb{E}R_1(x)^2+\omega\|\dot\eta(x_0)\|\right)^2 d\mathrm{Vol}^{d-1}(x_0)+r_5+o.
\end{aligned}$$

The remainder r_5 is not a small order term, but we can show that it is positive, and calculate its rate.

$$r_5 = O\left(\frac{B_1}{4k}+\int_S\frac{\|\dot\Psi(x_0)\|}{\|\dot\eta(x_0)\|^2}\left(b(x_0)\mathbb{E}R_1(x)^2+\omega\|\dot\eta(x_0)\|\right)^2 d\mathrm{Vol}^{d-1}(x_0)\right).$$

For r_5,

$$\begin{aligned}
r_5 \quad = \quad &\int_{\mathcal{S}}\int_0^{+\infty} t\|\dot\Psi(x_0)\|\Phi\left(-\frac{t\|\dot\eta(x_0)\|-\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0)}{\sqrt{s_{k,n}^2}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
&-\int_{\mathcal{S}}\int_0^{+\infty} t\|\dot\Psi(x_0)\|\Phi\left(-\frac{t\|\dot\eta(x_0)\|+\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0)}{\sqrt{s_{k,n}^2}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
= \quad &\int_{\mathcal{S}}\int_{-2\omega}^{+\infty} (t+2\omega)\|\dot\Psi(x_0)\|\Phi\left(-\frac{t\|\dot\eta(x_0)\|+\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0)}{\sqrt{s_{k,n}^2}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
&-\int_{\mathcal{S}}\int_0^{+\infty} t\|\dot\Psi(x_0)\|\Phi\left(-\frac{t\|\dot\eta(x_0)\|+\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0)}{\sqrt{s_{k,n}^2}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
= \quad &\int_{\mathcal{S}}\int_0^{\infty} 2\omega\|\dot\Psi(x_0)\|\Phi\left(-\frac{t\|\dot\eta(x_0)\|+\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0)}{\sqrt{s_{k,n}^2}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
&+\int_{\mathcal{S}}\int_{-2\omega}^{0} (t+2\omega)\|\dot\Psi(x_0)\|\Phi\left(-\frac{t\|\dot\eta(x_0)\|+\omega\|\dot\eta(x_0)\|}{\sqrt{s_{k,n}^2}}-\frac{b(x_0)t_{k,n}(x_0)}{\sqrt{s_{k,n}^2}}\right)dtd\mathrm{Vol}^{d-1}(x_0)\\
:= \quad &A+B.
\end{aligned}$$

From the format of A and B, we know that they are positive. When $t_{k,n}(x_0)$ and $1/\sqrt{k}$ both $\ll \omega$, A is an exponential tail (so we just ignore it) and for B we have:

$$B = \int_{\mathcal{S}} \|\dot{\Psi}(x_0)\| \omega^2 d\mathrm{Vol}^{d-1}(x_0) + O(\omega t_{k,n}(x_0) + \omega/\sqrt{k}).$$

□

D.4 Theorem 4

Proof of Theorem 4. First, it is easy to know that $\omega = O((1/n)^{1/d})$ since the nearest neighbor has an average distance of $O((1/n)^{1/d})$.

Second, there is a difference between pre-processed 1NN and random perturbation: in pre-processed 1NN, the nearest neighbor distributes approximately uniformly around x, while the other neighbors should have a distance to x larger than the nearest neighbor. However, this difference only affects the remainder term of regret, i.e., assuming whether or not the other neighbors are uniformly distributed in the ball $B(x, R_{k+1})$ does not affect our result.

As a result, taking expectation on the direction of δ,

$$\begin{aligned}
&\mathbb{E}\int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2}(b(x_0)t_{k,n}(x_0) + \delta^\top(x_0)\dot{\eta}(x_0))^2 d\mathrm{Vol}^{d-1}(x_0)\\
=&\int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2}\mathbb{E}(b(x_0)t_{k,n}(x_0) + \delta^\top(x_0)\dot{\eta}(x_0))^2 d\mathrm{Vol}^{d-1}(x_0)\\
=&\int_S \frac{\|\dot{\Psi}(x_0)\|}{\|\dot{\eta}(x_0)\|^2}\left(b^2(x_0)t_{k,n}^2(x_0)\right) d\mathrm{Vol}^{d-1}(x_0) + \Theta(\omega^2).
\end{aligned}$$

When $n^{-1/d} \gg n^{-2/(4+d)}$, i.e. $d > 4$, the dominant part of regret becomes $n^{-2/d}$. □

E Regret Convergence under General Smoothness Condition and Margin Condition

E.1 Model and Theorem

In this section, we will relax the conditions on the distribution of X and smoothness of η, and as a consequence, we only obtain the rate of the regret (without explicit form for the multiplicative constant). Technically, we will adopt the framework of Chaudhuri and Dasgupta (2014), and the following assumptions on the smoothness of η and the density of X are used instead of conditions [A.1]-[A.3].

B.1 Let λ be the Lebesgue measure on $\mathbb{R}^d$. There exists a positive pair (c_0, r_0) such that for any $x \in \mathcal{X}$,

$$\lambda(\mathcal{X} \cap B(x,r)) \geq c_0 \lambda(B(x,r)),$$

for any $0 < r \leq r_0$.

B.2 The support of X is compact.

B.3 Margin condition: $P(0 < |\eta(x) - 1/2| < t) \leq Bt^\beta$.

B.4 Smoothness of η: there exist some $\alpha > 0$ and $c_r > 0$, such that $|\eta(x+r) - \eta(x)| \leq \|r\|^\alpha$ for any x and $r \leq c_r$.

B.4' Smoothness of η: there exist some $\alpha > 0$ and $c_r > 0$, such that $|\eta(x+r) - \eta(x)| \leq \|r\|^\alpha$ for any x and $r \leq c_r$.

B.5 The density of X is finite and bounded away from 0.

Remark 1. *In Chaudhuri and Dasgupta (2014), the assumption of smoothness is made on $|\mathbb{E}(\eta(x')|x' \in B(x,r)) - \eta(x)|$, which is a weaker assumption compared with our B.4. However, under either random perturbation or adversarial attack, given a direction δ to obtain $\widetilde{x}$, the assumption in Chaudhuri and Dasgupta (2014) cannot be simply applied.*

The following theorem provide a general upper bound of regret for both perturbed and attacked data:

Theorem S. 4 (Convergence of Regret)**.** *Under [B.1] to [B.5], if for some $\delta > 0$, $k/n^{\delta} \to \infty$, taking*

$$k \asymp O(n^{2\alpha/(2\alpha+d)} \wedge (n^{2\alpha/d}\omega^{-2\alpha\beta})^{1/(2\alpha/d+\beta+1)}),$$

the regret becomes

$$Regret(n,\omega) = O\left(\omega^{\alpha(\beta+1)} \vee n^{-\alpha(\beta+1)/(2\alpha+d)}\right),$$

where $n^{-\alpha(\beta+1)/(2\alpha+d)}$ is the minimax rate of regret in k-NN.

Theorem 4 also reveals a sufficient condition when k-NN is consistent, i.e regret finally converges to 0: for both perturbed and attacked data, when $\omega = o(1)$, k-NN is still consistent using these two types of corrupted testing data.

Theorem S. 5 (Minimax Rate of Regret)**.** *Let $\widehat{g}_n$ be an estimator of g, let $\mathcal{P}_{\alpha,\beta}$ be a set of distributions which satisfy [B.1] to [B.3], [B.4'], and [B.5], when $\alpha \leq 1$, there exists some $C > 0$ such that*

$$\sup_{P\in\mathcal{P}_{\alpha,\beta}} P(\widehat{g}_n(\widetilde{X}) \neq Y) - P(g(X) \neq Y) \geq C(\omega^{\alpha(\beta+1)} \vee n^{-\frac{\alpha(\beta+1)}{2\alpha+d}}). \tag{16}$$

The constant C depends on α, β, d only.

Theorem 5 reveals that, for any estimator of g, under either random perturbation or adversarial attack, the regret in the worst case is larger than $C(\omega^{\alpha(\beta+1)} \vee n^{-\frac{\alpha(\beta+1)}{2\alpha+d}})$. Theorem 4 and 5 together shows that the kNN estimator reaches the optimal rate of regret.

E.2 Proofs

Proof of Theorem S.4. Let $p = k/n$. Denote $R_{k,n}(x) = P(\widehat{g}_{k,n}(x) \neq Y|x)$ and $R^*(x) = P(g(x) \neq Y)$, and $\mathbb{E}R_{k,n}(x) - R^*(x)$ as the excess risk. Define

$$\mathcal{X}^+_{p,\Delta,\omega} = \{x \in \mathcal{X} | \eta(x) > \frac{1}{2}, \forall x' \in B(x,\omega), \eta(x'+r) \geq \frac{1}{2} + \Delta, \forall \|r\| < r_{2p}(x)\},$$
$$\mathcal{X}^-_{p,\Delta,\omega} = \{x \in \mathcal{X} | \eta(x) < \frac{1}{2}, \forall x' \in B(x,\omega), \eta(x'+r) \leq \frac{1}{2} - \Delta, \forall \|r\| < r_{2p}(x)\},$$

with r_{2p} as the distance from x to its $2pn$th nearest neighbor, and the decision boundary area:

$$\partial_{p,\Delta,\omega} = \mathcal{X} \setminus (\mathcal{X}^+_{p,\Delta,\omega} \cup \mathcal{X}^-_{p,\Delta,\omega}).$$

Given $\partial_{p,\Delta,\omega}$, $\mathcal{X}^+_{p,\Delta,\omega}$, and $\mathcal{X}^-_{p,\Delta,\omega}$, similar with Lemma 8 in Chaudhuri and Dasgupta (2014), the event of $g(x) \neq \widehat{g}_{k,n}(x)$ can be covered as:

$$\begin{aligned} 1_{\{g(x)\neq\widehat{g}_{k,n}(x)\}} \leq\ & 1_{\{x\in\partial_{p,\Delta,\omega}\}} \\ & +1_{\{\max_{i=1,\ldots,k} R_i(\widetilde{x})\geq r_{2p}(x)\}} \\ & +1_{\{|\widehat{\eta}_{k,n}(x)-\eta(x'+r)|\geq\Delta\}}. \end{aligned}$$

When $\eta(x'+r) > 1/2$ for all $\|r\| \leq r_{2p}(x)$, and $x \in \mathcal{X}^+_{p,\Delta}$, assume $\widehat{\eta}_{k,n}(x) < 1/2$, then

$$\eta(x'+r) - \widehat{\eta}_{k,n}(x') > \eta(x'+r) - 1/2 \geq \Delta.$$

The other two events are easy to figure out.

By Chaudhuri and Dasgupta (2014) and Belkin et al. (2018), $P(\max_{i=1,\ldots,k} R_i(x) \geq r_{2p}(x))$ is of $O(\exp(-ck^2))$ for some $c > 0$, hence it becomes a smaller order term if for some $\delta > 0$, $k/n^{\delta} \to \infty$.

In addition, from the definition of regret, assume $\eta(x) < 1/2$,

$$\begin{aligned}
& P(\widehat{g}(x) \neq Y|X=x) - \eta(x) \\
= \;& \eta(x) P(\widehat{g}(x) = 0|X=x) + (1-\eta(x))P(\widehat{g}(x)=1|X=x) - \eta(x) \\
= \;& \eta(x) P(\widehat{g}(x) = g(x)|X=x) + (1-\eta(x))P(\widehat{g}(x) \neq g(x)|X=x) - \eta(x) \\
= \;& \eta(x) - \eta(x) P(\widehat{g}(x) \neq g(x)|X=x) + (1-\eta(x))P(\widehat{g}(x) \neq g(x)|X=x) - \eta(x) \\
= \;& (1-2\eta(x))P(\widehat{g}(x) \neq g(x)|X=x),
\end{aligned}$$

similarly, when $\eta(x) > 1/2$, we have

$$P(\widehat{g}(x) \neq Y|X=x) - 1 + \eta(x) \;\; = \;\; (2\eta(x)-1)P(\widehat{g}(x) \neq g(x)|X=x).$$

As a result, the regret can be represented as

$$Regret(k,n,\omega) \;\; = \;\; \mathbb{E}\left(|1-2\eta(X)|P(g(X) \neq \widehat{g}_{k,n}(X))\right).$$

For simplicity, denote $p = k/n$. We then follow the proof of Lemma 20 of Chaudhuri and Dasgupta (2014). Without loss of generality assume $\eta(x) > 1/2$. For perturbation $\delta \in \mathbb{R}^d$, define

$$\begin{aligned}
\Delta_0 &= \sup_{x,\delta,\|r\|<r_{2p}(x)} |\eta(x+\delta+r) - \eta(x)| = O(\omega^\alpha) + O((k/n)^{\alpha/d}), \\
\Delta(x) &= |\eta(x) - 1/2|,
\end{aligned}$$

then we have

$$\eta(x+\delta+r) \geq \eta(x) - \Delta_0 = \frac{1}{2} + (\Delta(x) - \Delta_0),$$

hence $x \in \mathcal{X}^+_{p,\Delta(x)-\Delta_0,\omega}$.

From the definition of $R_{k,n}$ and R^*, when $\Delta(x) > \Delta_0$, we also have

$$\begin{aligned}
& \mathbb{E}R_{k,n}(x) - R^*(x) \\
\leq \;& 2\Delta(x)\bigg[P(r_{(k+1)} > v_{2p}) + P\bigg(\sum_{i=1}^k \frac{1}{k}Y(X_i) - \eta(x'+\delta+r) > \Delta(x) - \Delta_0\bigg)\bigg] \\
\leq \;& 2\Delta(x) P\bigg(\sum_{i=1}^k \frac{1}{k}Y(X_i) - \eta(x'+\delta+r) > \Delta(x) - \Delta_0\bigg) + o \\
= \;& 2\Delta(x)\mathbb{E}_\delta\left[P\bigg(\sum_{i=1}^k \frac{1}{k}Y(X_i) - \eta(x'+\delta+r) > \Delta(x) - \Delta_0 \bigg| \delta\bigg)\right] + o
\end{aligned}$$

Considering the problem that the upper bound can be much greater than 1 when $\Delta(x)$ is small, we define $\Delta_i = 2^i\Delta_0$, taking $i_0 = \min\{i \geq 1 |\; (\Delta_i - \Delta_0)^2 > 1/k\}$, using Berstein inequality, it becomes

$$\begin{aligned}
\mathbb{E}R_{k,n}(X) - R^*(X) \;\; &= \;\; \mathbb{E}(R_{k,n}(X) - R^*(X))1_{\{\Delta(X) \leq \Delta_{i_0}\}} \\
&\quad\;\; + \mathbb{E}(R_{k,n}(X) - R^*(X))1_{\{\Delta(X) > \Delta_{i_0}\}} \\
&\leq \;\; 2\Delta_{i_0} P(\Delta(X) \leq \Delta_{i_0}) + \exp(-k/8) \\
&\quad\;\; + c_2\mathbb{E}\left[\Delta(X)1_{\{\Delta_{i_0} < \Delta(X)\}} \exp(-c_1 k(\Delta(x) - \Delta_0)^2)\right] \\
&\leq \;\; 2\Delta_{i_0} P(\Delta(X) \leq \Delta_{i_0}) + \exp(-k/8) \\
&\quad\;\; + c_2\mathbb{E}\left[\Delta(X)1_{\{\Delta_{i_0} < \Delta(X)\}} \exp(-c_1 k(\Delta(x) - \Delta_0)^2)\right].
\end{aligned}$$

When $i_0 = \min\{i \geq 1 | (\Delta_i - \Delta_0)^2 > 1/k\}$, the exponential tail will diminish fast, leading to

$$\begin{aligned} & \mathbb{E}\left[\Delta(X)1_{\{\Delta_{i_0}<\Delta(X)\}} \exp(-c_1 k(\Delta(x)-\Delta_0)^2)\right] \\ = & \sum_{i=i_0}^{\infty} \mathbb{E}\left[\Delta(X)1_{\{\Delta_i<\Delta(X)<\Delta_{i+1}\}} \exp(-c_1 k(\Delta(x)-\Delta_0)^2)\right] \\ \leq & \sum_{i=i_0}^{\infty} \Delta_{i+1}^{\beta+1} \exp(-c_1 k(\Delta_i-\Delta_0)^2) \\ = & \sum_{i=i_0}^{\infty} \Delta_0^{\beta+1} 2^{(i+1)(\beta+1)} \exp(-c_1 k\Delta_0^2(2^i-1)^2) \\ \leq & c_3 \Delta_0^{\beta+1}. \end{aligned}$$

Recall that $\Delta_{i_0} > \Delta_0$ and $\Delta_{i_0}^2 > 1/k$, hence when $\Delta_{i_0}^2 = O(1/k)$, we can obtain the minimum upper bound

$$\mathbb{E}R_{k,n}(X) - R^*(X) = O(\Delta_0^{\beta+1}) + O\left(\left(\frac{1}{k}\right)^{(\beta+1)/2}\right).$$

□

Proof of Theorem S.5. The proof is similar as Audibert and Tsybakov (2007) using technical details in Audibert (2004) for Assouad's method. There are two scenarios we will consider. Define C_0, C_1 and C_2 as some suitable constants, we will first show for any $\omega \geq 0$,

$$\sup_{P\in\mathcal{P}_{\alpha,\beta}} P(\widehat{g}_n(\widetilde{X}) \neq Y) - P(g(X) \neq Y) \geq C_1 n^{-\frac{\alpha(\beta+1)}{2\alpha+d}}. \tag{17}$$

Further, when $\omega > C_0 n^{-\frac{1}{2\alpha+d}}$, our target is to show that

$$\sup_{P\in\mathcal{P}_{\alpha,\beta}} P(\widehat{g}_n(\widetilde{X}) \neq Y) - P(g(X) \neq Y) \geq C_2 \omega^{\alpha(\beta+1)}. \tag{18}$$

Case 1: when $\omega \leq C_0 n^{-\frac{1}{2\alpha+d}}$, the basic idea is to construct a distribution of x and two distributions of $y|x$ such that, the Bayes classifiers from these two distributions of $y|x$ reverse with each other, but through sampling n points, we cannot distinguish which distribution these n samples chosen are from. For example, given n samples from a normal distribution, statistically we cannot determine whether data are sampled from a zero-mean distribution, or a distribution with mean $1/\sqrt{n}$, thus any estimator based on data (either using clean testing data or corrupted testing data) can make a false prediction.

Assume X distributed within a compact set in $[0,1]^d$. For an integer $q \geq 1$, consider the regular grid as

$$G_q := \left\{\left(\frac{2k_1+1}{2q}, ..., \frac{2k_d+1}{2q}\right) : k_i \in \{0, ..., q-1\}, i = 1, ..., d\right\}. \tag{19}$$

For any point x, denote $n_q(x)$ as the closest grid point in G_q, and define $\mathcal{X}_1', ..., \mathcal{X}_{q^d}'$ as a partition of $[0,1]^d$ such that x and x' are in the same $\mathcal{X}_i'$ if and only if $n_q(x) = n_q(x')$. Among all the $\mathcal{X}_i'$'s, select m of them as $\mathcal{X}_1, ..., \mathcal{X}_m$, and $\mathcal{X}_0 := [0,1]^d \backslash \cup_{i=1}^m \mathcal{X}_i$.

Take z_i as the center of $\mathcal{X}_i$ for $i = 1, ..., m$. When $x \in B(z_i, 1/4q)$, set the density of x as $\epsilon/\lambda[B(z_i, 1/4q)]$ for some $\epsilon > 0$, and the density of x in $\mathcal{X}_i \backslash B(z_i, 1/4q)$ is set to be 0. Assume x uniformly distributes in $\mathcal{X}_0$.

Let $u : \mathbb{R}^+ \to \mathbb{R}^+$ be a nonincreasing infinitely differentiable function starting from 0 and satisfying α-smoothness condition. Moreover, u is 1 in $[1/2, \infty)$. Denote ψ and ϕ as

$$\psi(x) := C_\psi u(\|x\|), \tag{20}$$

and

$$\phi(x) := q^{-\alpha}\psi(q(x - n_q(x))). \tag{21}$$

Through the above construction, if we take $\eta(x)=(1+\phi(x))/2$ or $\eta(x)=(1-\phi(x))/2$, and let $m=O(q^{d-\alpha\beta})$, then when $\alpha\beta\leq d$, β margin condition is also satisfied.

The construction above will also be applied in *Case 2* (with difference on the choice of q,ϵ,u).

Now we apply Assouad's method to find the lower bound of regret. Denote P_{jk} as a distribution such that $\eta(x)=(1+\phi(x))/2$ when $k=0$, $x\in\mathcal{X}_j$, and $\eta(x)=(1-\phi(x))/2$ when $k=1$, $x\in\mathcal{X}_j$, then we have for any estimator $\widehat{g}(x,Z_n)$ with $Z_n=(X_n,Y_n)$ as data,

$$\sup_{k=0,1}\mathbb{E}_{X,Z_n,P_{jk}}1_{\{\widehat{g}(X,Z_n)\neq g(X)\}}1_{\{X\in\mathcal{X}_j\}} \tag{22}$$

$$\geq \frac{1}{2}\mathbb{E}_{X,Z_n,P_{j0}}1_{\{\widehat{g}(X,Z_n)\neq g(X)\}}1_{\{X\in\mathcal{X}_j\}}+\frac{1}{2}\mathbb{E}_{Z_n,P_{j1}}1_{\{\widehat{g}(X,Z_n)\neq g(X)\}}1_{\{X\in\mathcal{X}_j\}} \tag{23}$$

$$= \frac{1}{2}\mathbb{E}_{X,Z_n,P_{j0}}1_{\{\widehat{g}(X,Z_n)\neq 0\}}1_{\{X\in\mathcal{X}_j\}}+\frac{1}{2}\mathbb{E}_{Z_n,P_{j1}}1_{\{\widehat{g}(X,Z_n)\neq 1\}}1_{\{X\in\mathcal{X}_j\}} \tag{24}$$

$$= \frac{1}{2}\mathbb{E}_X 1_{\{X\in\mathcal{X}_j\}}\mathbb{E}\left[\mathbb{E}_{Z_n,P_{j0}}1_{\{\widehat{g}(x,Z_n)\neq 0\}}+\mathbb{E}_{Z_n,P_{j1}}1_{\{\widehat{g}(x,Z_n)\neq 1\}}\middle|X=x\right] \tag{25}$$

$$= \frac{1}{2}\mathbb{E}_X 1_{\{X\in\mathcal{X}_j\}}\mathbb{E}\left[\int 1_{\{\widehat{g}(x,Z_n)\neq 0\}}dP_{j0}(Z_n)+\int 1_{\{\widehat{g}(x,Z_n)\neq 1\}}dP_{j1}(Z_n)\middle|X=x\right] \tag{26}$$

$$\geq \frac{1}{2}\mathbb{E}_X 1_{\{X\in\mathcal{X}_j\}}\mathbb{E}\left[\int 1_{\{\widehat{g}(x,Z_n)\neq 0\}}+1_{\{\widehat{g}(x,Z_n)\neq 1\}}(dP_{j0}(Z_n)\wedge dP_{j1}(Z_n))\middle|X=x\right] \tag{27}$$

$$= \frac{1}{2}\mathbb{E}_X 1_{\{X\in\mathcal{X}_j\}}\mathbb{E}\left[\int (dP_{j0}(Z_n)\wedge dP_{j1}(Z_n))\middle|X=x\right] \tag{28}$$

$$= \frac{1}{2}\mathbb{E}_X 1_{\{X\in\mathcal{X}_j\}}\int (dP_{j0}(Z_n)\wedge dP_{j1}(Z_n)). \tag{29}$$

Denote

$$b_j:=\left[1-\mathbb{E}^2(\sqrt{1-\phi^2(X)}|X\in\mathcal{X}_j)\right]^{1/2}, \tag{30}$$

and

$$b_j':=(\mathbb{E}\phi(X)|X\in\mathcal{X}_j), \tag{31}$$

then $\int(dP_{j0}(Z_n)\wedge dP_{j1}(Z_n))=\Theta(1)$ through our design of $\mathcal{X}_j$ when $b_j=O(1/\sqrt{n\epsilon})$ by Lemma 5.1 in Audibert (2004).

As a result, when $b_j=b$, $b_j'=b'$ for all $j=1,...,m$, and $b=O(1/\sqrt{n\epsilon})$, there exists some $C_3>0$ such that

$$\sup_{P\in\mathcal{P}}P(\widehat{g}(X,Z_n)\neq Y)-P(g(X)\neq Y) \tag{32}$$

$$= \sup_{P\in\mathcal{P}}\mathbb{E}|2\eta(X)-1|P(\widehat{g}(X,Z_n)\neq g(X)) \tag{33}$$

$$= \sup_{P\in\mathcal{P}}\sum_{j=1}^{m}\mathbb{E}|2\eta(X)-1|P(\widehat{g}(X,Z_n)\neq g(X))1_{\{X\in\mathcal{X}_j\}} \tag{34}$$

$$\geq C_3 m b'\epsilon. \tag{35}$$

The regret is lower bounded as $C_1 n^{-\alpha(\beta+1)/(2\alpha+d)}$ when taking $q=O(n^{1/(2\alpha+d)})$. Note that $\widehat{g}(x,Z_n)$ can be any classifier, which also includes those "random" estimators when x is perturbed / attacked.

Case 2: when $\omega>C_0 n^{-\frac{1}{2\alpha+d}}$, we construct a distribution of (x,y) such that, after injecting noise in it, there is some sets of $\tilde{x}$ where $P(g(x)=1|\tilde{x})$ and $P(g(x)=0|\tilde{x})$ are comparable, thus no matter which label is obtained from the estimator, it has a constant-level of probability to make false decision at this $\tilde{x}$.

The construction is similar as *Case 1*, and we take $q=\lfloor 2/\omega\rfloor$. For function u, here we let it increase from 0 and becomes 1 in $[1/4,\infty)$. For each pair $(\mathcal{X}_{j0},\mathcal{X}_{j1})$, take $\eta(x)=(1+\phi(x))/2$ when $x\in\mathcal{X}_{j0}$ and $\eta(x)=(1-\phi(x))/2$ when $x\in\mathcal{X}_{j1}$. The support of x is $\mathcal{X}_0\cup(\bigcup_{i=1}^m B(z_i,3\omega/4))$. Take $m=O(\omega^{\alpha\beta-d})$ and $\epsilon=O(\omega^d)$, then both α-smoothness condition and β-margin condition are satisfied.

After injecting random noise on x, consider ξ_j as the boundary between $\mathcal{X}_{j0}$ and $\mathcal{X}_{j1}$, then when $\tilde{x}$ is from $\{z \mid dist(z, \xi_j) < \omega/4,\ z \in \mathcal{X}_{j0} \cup \mathcal{X}_{j1}\}$, $P(g(x) = 1|\tilde{x})$ and $P(g(x) = 0|\tilde{x})$ are in $[C_4, 1 - C_4]$ for some constant $C_4 > 0$. Thus the probability of any estimator to make a false decision at this $\tilde{x}$ is larger than C_4. In addition, the probability measure of $\cup_{j=1}^{m}\{z \mid dist(z, \xi_j) < \omega/4,\ z \in \mathcal{X}_{j0} \cup \mathcal{X}_{j1}\}$ is greater than $C_5\omega^{\alpha\beta}$ for some constant $C_5 > 0$. Thus the regret is greater than $C_5\omega^{\alpha\beta}C_\phi\omega^{\alpha}C_4 = C_6\omega^{\alpha(\beta+1)}$.

□

References

Audibert, J.-Y. (2004), “Classification under polynomial entropy and margin assumptions and randomized estimators,” .

Audibert, J.-Y. and Tsybakov, A. B. (2007), “Fast learning rates for plug-in classifiers,” *The Annals of statistics*, 35, 608–633.

Belkin, M., Hsu, D., and Mitra, P. (2018), “Overfitting or perfect fitting? Risk bounds for classification and regression rules that interpolate,” *arXiv preprint arXiv:1806.05161*.

Chaudhuri, K. and Dasgupta, S. (2014), “Rates of convergence for nearest neighbor classification,” in *Advances in Neural Information Processing Systems*, pp. 3437–3445.

Gray, A. (2012), *Tubes*, vol. 221, Birkhäuser.

Papernot, N. and McDaniel, P. (2018), “Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning,” *arXiv preprint arXiv:1803.04765*.

Samworth, R. J. (2012), “Optimal weighted nearest neighbour classifiers,” *The Annals of Statistics*, 40, 2733–2763.

Sun, W. W., Qiao, X., and Cheng, G. (2016), “Stabilized nearest neighbor classifier and its statistical properties,” *Journal of the American Statistical Association*, 111, 1254–1265.

Wang, Y., Jha, S., and Chaudhuri, K. (2017), “Analyzing the robustness of nearest neighbors to adversarial examples,” *arXiv preprint arXiv:1706.03922*.

Xing, Y., Song, Q., and Cheng, G. (2018), “Statistical optimality of interpolated nearest neighbor algorithms,” *arXiv preprint arXiv:1810.02814*.