STAT 583: Advanced Theory of Statistical Inference **Homework 3**

Lecture 3: Homework

Lecturer: Fang Han *Due Date: Jun. 1*

- Please send the electronic copies of your HW to Yandi Shen's email box (*ydshen@uw.edu*).

3.1 VC-stability

Problem 1

Prove Lemma 32 in Chapter 2 (Lemma 9.14 in K2008).

3.2 Covering number

Problem 2

Prove, if $\mathcal{F}$ is compact, then the function $\epsilon \to N(\mathcal{F}, L_2(P), \epsilon)$ is left-continuous (it is covering with open balls).

3.3 M-estimation and Z-estimation

Problem 3

Following Example 6 in Chapter 3 (Example 3.2.23 in VW1996), please formulate a rigorous statement for the ASN of the quantile regression estimator. In detail, you should lay out sufficient conditions under which $\widehat{\theta}_n$ is ASN, calculate the asymptotic covariance of $\widehat{\theta}_n$, and prove your result.

Problem 4

Prove Corollary 10 in Chapter 3: please formulate a rigorous statement for the ASN of the Z-estimator satisfying the Lipschitz condition. In detail, you should lay out sufficient conditions under which $\widehat{\theta}_n$ is ASN, calculate the asymptotic covariance of $\widehat{\theta}_n$, and prove your result.

3.4 Tail bounds

In the following, you might consider using the following V-C theorem.

Theorem (Vapnik and Chervonenkis (1971)). For any probability measure P and class of sets $\mathcal{A}$, and for any n and $\epsilon > 0$,

$$P\Big(\sup_{A\in\mathcal{A}} |\mathbb{P}_n(A) - P(A)| > \epsilon\Big) \leq 8\Pi_{\mathcal{A}}(n)\exp(-n\epsilon^2/32),$$

where $\mathbb{P}_n$ stands for the empirical measure.

Problem 5 (k-means clustering)

Let $X_1, \ldots, X_n$ be n i.i.d. random variables from a distribution supported on $[-B, B]^d$ for some $B < \infty$. The k-means method of clustering is based on choosing a set of k vectors $a_1, \ldots, a_k$ in $\mathbb{R}^d$ (representing centers of k clusters) to minimize the empirical risk:

$$\widehat{M}_n(a_1, \ldots, a_k) := \frac{1}{n}\sum_{i=1}^{n} \min_j \|X_i - a_j\|_2^2.$$

The population error of the clustering can be measured by the quantity:

$$M(a_1, \ldots, a_k) = \mathbb{E}[\min_j \|X - a_j\|_2^2 \mid X_1, \ldots, X_n]$$

where X is an independent draw from the same distribution. Note that $M(a_1, \ldots, a_k)$ is still a random variable w.r.t. $X_1, \ldots, X_n$.

a. If $(a_1, \ldots, a_k)$ are a set of empirically optimal cluster centers (a.k.a., the ERM) satisfying

$$\widehat{M}_n(a_1, \ldots, a_k) = \inf_{b_1,\ldots,b_k} \widehat{M}_n(b_1, \ldots, b_k),$$

show that

$$M(a_1, \ldots, a_k) - \inf_{b_1,\ldots,b_k\in\mathbb{R}^d} M(b_1, \ldots, b_k) \leq 2 \sup_{b_1,\ldots,b_k\in\mathbb{R}^d} |\widehat{M}_n(b_1, \ldots, b_k) - M(b_1, \ldots, b_k)|.$$

b. Define $f_b(x) := \min_{j=1,\ldots,k} \|x - b_j\|_2^2$. Prove that $0 \leq f_b(x) \leq 4dB^2$.

c. Prove that

$$\sup_{b\in(\mathbb{R}^d)^k} \left|\widehat{M}_n(b) - M(b)\right| \leq 4dB^2 \sup_{b\in(\mathbb{R}^d)^k, 0\leq t\leq 4dB^2} \left|\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(f_b(X_i) > t) - \mathbb{P}(f_b(X) > t)\right|.$$

d. Show that, for all $\epsilon > 0$,

$$\mathbb{P}\left\{\sup_{b_1,\ldots,b_k\in\mathbb{R}^d} |\widehat{M}_n(b_1, \ldots, b_k) - M(b_1, \ldots, b_k)| > \epsilon\right\} \leq C(n+1)^{k(d+2)}\exp\left(-\frac{n\epsilon^2}{C'd^2B^4}\right),$$

for some absolute constants C, C' independent of (n, k, d, B).

e. Conclude that the population error of the ERM converges in probability to $\inf_{b_1,\ldots,b_k\in\mathbb{R}^d} M(b_1, \ldots, b_k)$ as $n \to \infty$ and k, d fixed.

Problem 6 (Generalized Glivenko-Cantelli Theorem)

Let $\mathcal{F}$ be a class of bounded functions satisfying $0 \leq f(x) \leq M$ for all $x \in \mathbb{R}^d$. Define the class of sets

$$\mathcal{A} = \{A_{f,t} : f \in \mathcal{F}, t \in [0, M]\},$$

where for every $f \in \mathcal{F}$ and $t \in [0, M]$, the set $A_{f,t} \in \mathbb{R}^d$ is defined as

$$A_{f,t} = \{z : f(z) > t\}.$$

Prove

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \mathbb{E} f(X_i) \right| > \epsilon \right) \leq 8 \Pi_{\mathcal{A}}(n) e^{-n\epsilon^2/(32M^2)}.$$

Problem 7

Use the result of Problem 5 or 6 to reconstruct a proof of the VC-major theorem: Suppose $\mathcal{F}$ is a bounded VC-major class. Prove

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \overset{a.s.}{\to} 0.$$

Please also work out the rate of convergence (needs not be the sharpest).

3.5 Applications

Problem 8

Redo Problems 7 and 8 in HW1, but this time without assuming any Donsker property on any function class.