36-755: Advanced Statistical Theory I
Final Exam
December, 7, 2016

Instructions:

- **Duration: 1 hour and 20 minutes.**
- **This is an open-notes, open-books exam. You may use your laptop as long as you are not connected to the internet.**
- **There are 7 problems, each worth 25 points. Your score will be capped at 100.**
- **YOU ARE NOT REQUIRED TO CARRY OUT ALL THE CALCULATIONS. To receive full credit, it will be enough to set them up correctly and to indicate which results/tools you are using. You do not need to be concerned with providing exact constants.**

1. Assume that X is a vector in $\mathbb{R}^d$ that is sub-Gaussian with parameter σ^2 (this means that $v^\top X \in SG(\sigma^2)$ for each $v \in \mathbb{R}^d$ with $\|v\| = 1$). Compute upper bounds for

$$\mathbb{P}\left(\|X\| \geq t\right), \quad t > 0,$$

and

$$\mathbb{E}\left[\|X\|\right].$$

Hint: Use the fact that, for any $x \in \mathbb{R}^d$, $\|x\| = \max_{\{v \in \mathbb{R}^d, \|v\| \leq 1\}} v^\top x$. Also, recall that the δ-covering number of the Euclidean unit ball in $\mathbb{R}^d$ is bounded by $\left(1 + \frac{2}{\delta}\right)^d$.

2. Let $\mathcal{F}_{\alpha,\gamma}(C_{\max}, L)$ denotes the class of real-valued functions from $[0,1]$ such that, for $\gamma \in (0,1]$, $\alpha \in \mathbb{N}$, $C_{\max} > 0$ and $L > 0$,

$$\sup_{x \in [0,1]} |f^{(j)}(x)| \leq C_{\max}, \quad j = 0, 1, \ldots, \alpha$$

and

$$|f^{(\alpha)}(x) - f^{(\alpha)}(y)| \leq L|x - y|^\gamma,$$

where $f^{(k)}$ denotes the kth order derivative of f. It is a well-known fact that, for some C depending on L, α, γ and $C_{\max}$,

$$\log N(\delta, \mathcal{F}_{\alpha,\gamma}(C_{\max}, L), \|\cdot\|_\infty) \leq C \left(\frac{1}{\delta}\right)^{\frac{1}{\alpha+\gamma}},$$

where $N(\delta, \mathcal{F}_{\alpha,\gamma}(C_{\max}, L), \|\cdot\|_\infty)$ is the δ-covering number of $\mathcal{F}_{\alpha,\gamma}(C_{\max}, L)$ in the distance induced by the $\|\cdot\|_\infty$ norm, where $\|f\|_\infty = \sup_{x \in [0,1]} |f(x)|$.

Suppose we observe

$$Y_i = f^*(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where $f^* \in \mathcal{F}_{\alpha,\gamma}(C_{\max}, L)$, the ϵ_i's are i.i.d. standard Gaussian and the x_i's are deterministic points in $[0,1]$.

Consider the non-parametric least-squares estimator

$$\hat{f} \in \operatorname{argmin}_{f \in \mathcal{F}_{\alpha,\gamma}(C_{\max},L)} \frac{1}{n}\sum_{i=1}^{n}(Y_i - f(x_i))^2 .$$

Explain how you can compute, using arguments based on the notion of local Gaussian complexity, a high-probability bound for

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{f}(x_i) - f^*(x_i)\right)^2 .$$

You do not need to carry out all the calculations to determine the bound: it is enough to set them up.

3. Consider the same settings as in the previous problem. Derive the basic inequality

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{f}(x_i) - f^*(x_i)\right)^2 \leq \frac{2}{n}\sum_{i=1}^{n}\epsilon_i(\widehat{f}(x_i) - f^*(x_i)).$$

Starting from this inequality, explain how an application of the naive 1-step discretization bound will yield a bound on

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{f}(x_i) - f^*(x_i)\right)^2\right].$$

For the second part of the question, you do not need to carry out all the calculation to determine the bound: it is enough to set them up.

4. Let $X_1, \ldots, X_n$ be i.i.d. samples from a probability distribution over the real line with Lebesgue density f. A standard estimator of f is the *kernel density estimator*

$$\widehat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \quad x \in \mathbb{R},$$

where $K : \mathbb{R} \to [0, \infty)$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(u)du = 1$ and $h > 0$ is fixed bandwidth parameter. We assess the quality of the estimator $\widehat{f}_h$ using its L_1 distance from f:

$$\|\widehat{f}_h - f\|_1 := \int_{-\infty}^{\infty} |\widehat{f}_h(u) - f(u)|du$$

Prove that $\|\widehat{f}_h - f\|_1$ concentrates well around its mean, i.e. derive an exponential bound for the probability

$$\mathbb{P}\left(\left|\|\widehat{f}_h - f\|_1 - \mathbb{E}\|\widehat{f}_h - f\|_1\right| \geq t\right),$$

for any $t > 0$.

5. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a probability distribution on $\mathbb{R}^d$. Let F denote the multivariate c.d.f. of P, i.e.

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}^d,$$

where, for vectors x and y in $\mathbb{R}^d$, $x \le y$ means that $x_j \le y_j$ for all $j = 1, \ldots, d$. Let $\widehat{F}_n$ denote the empirical c.d.f. of P, i.e.

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \le x), \quad x \in \mathbb{R}^d.$$

Derive a suitable bound for the probability

$$\mathbb{P}\left(\|F - \widehat{F}_n\|_\infty > t\right), \quad \forall t > 0,$$

where $\|F - \widehat{F}_n\|_\infty = \sup_{x\in\mathbb{R}^d} |F(x) - \widehat{F}_n(x)|$. You may use the fact that the VC dimension of the class of sets

$$\mathcal{A} = \left\{(-\infty, x_1] \times \ldots \times (-\infty, x_d], \ \ (x_1, \ldots, x_d) \in \mathbb{R}^d\right\}$$

is d.

Use the above result to derive a $1-\alpha$ confidence set for F, where $\alpha \in (0, 1)$.

6. Let $X_1, \ldots, X_n$ an i.i.d. sample from a probability distribution P with mean μ and variance σ^2. Suppose we want to estimate μ^2 using the U-statistic

$$U_n = \binom{n}{2}^{-1} \sum_{i<j} X_i X_j.$$

Show that the asymptotic distribution of $\sqrt{n}(U_n - \mu^2)$ is $N(0, 4\mu^2\sigma^2)$.

What happens when $\mu = 0$?

Using the fact that we can rewrite U_n as

$$\frac{1}{n-1}\left(\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right)$$

show that, when $\mu = 0$, nU_n has asymptotically the distribution of $(Z^2 - 1)\sigma^2$, where $Z \sim N(0, 1)$.

7. (**Spiked covariance model**). Let $X_1, \ldots, X_n$ be an *i.i.d.* sample from a probability distribution on $\mathbb{R}^d$ with mean zero and covariance

$$\Sigma = \theta v v^\top + I_d,$$

where $\theta > 0$ and $\|v\| = 1$. Then, for all $i = 1, \ldots, n$,

$$X_i \overset{d}{=} \sqrt{\theta}\xi_i v + \epsilon_i$$

where $\overset{d}{=}$ denotes equality in distribution, $(\xi_1, \ldots, \xi_n)$ are independent zero mean variables with unit variance and $(\epsilon_1, \ldots, \epsilon_n)$ are independent vectors (independent of the ξ's) with mean zero and common covariance I_d. Assume that the ξ's and the ϵ_i's are sub-Gaussian variates with parameters at most 1.

We saw in class that in order to analyze the performance of PCA, we need to control $\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}}$, where $\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} X_i X_i^\top$.

Show that

$$\|\widehat{\Sigma} - \Sigma\|_{\mathrm{op}} \le T_1 + T_2 + T_3,$$

where

$$T_1 = \theta \left| \frac{1}{n} \sum_{i=1}^{n} \xi_i^2 - 1 \right|,$$

$$T_2 = 2\sqrt{\theta} \left\| \frac{1}{n} \sum_{i=1}^{n} \xi_i \epsilon_i \right\|$$

and

$$T_3 = \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \epsilon_i^\top - I_d \right\|_{\mathrm{op}}.$$

Explain how to obtain high-probability bounds for each of the terms T_1, T_2 and T_3. (*For the term T_2 see the hint in problem 1...*)
For the second part of the question, you do not need to carry out any calculations: it is enough to explain which tools you would use for each term.