1. (Linear regression) Consider the model

\[ Y = X \beta + Z \]

where the design matrix \( X \in \mathbb{R}^{n \times d} \) is known and \( Z \sim N(0, I_n) \). Define the minimax mean-square error of estimating the regression coefficient \( \beta \in \mathbb{R}^d \) based on \( X \) and \( Y \) as follows:

\[ R^{*}_{\text{est}} = \inf_{\hat{\beta}} \sup_{\beta} \mathbb{E} \| \hat{\beta} - \beta \|_2^2. \]  

This setup is very general in that one can define \( X \) appropriately to handle the cases of multiple observations, general covariance, etc.

(a) Show that if \( \text{rank}(X) < d \), then \( R^{*}_{\text{est}} = \infty \);
(b) Show that if \( \text{rank}(X) = d \), then

\[ R^{*}_{\text{est}} = \text{tr}((X^\top X)^{-1}) \]

and identify which estimator achieves the minimax risk.
(c) As opposed to the estimation error in (1), consider the prediction error:

\[ R^{*}_{\text{pred}} = \inf_{\hat{\beta}} \sup_{\beta} \mathbb{E} \| X \hat{\beta} - X \beta \|_2^2. \]  

(2)

Redo (a) and (b) by finding the value of \( R^{*}_{\text{pred}} \) and identify the minimax estimator. Explain intuitively why \( R^{*}_{\text{pred}} \) is always finite even when \( d \) exceeds \( n \).

2. (Sharp minimax rate in sparse denoising) For \( d \)-dimensional GLM model \( X \sim N(\theta, I_d) \), we show that minimax risk for denoising \( k \)-sparse vectors in high dimensions: as \( d \to \infty \) and \( k/d \to 0 \),

\[ R^{*}(k, d) \triangleq \inf_{\hat{\theta}} \sup_{\|\theta\|_0 \leq k} \mathbb{E}_{\theta}[\| \hat{\theta} - \theta \|_2^2] = (2 + o(1))k \log \frac{d}{k}. \]  

(3)

(a) We first consider 1-sparse vectors and prove

\[ R^{*}(1, d) \triangleq \inf_{\theta} \sup_{\|\theta\|_0 \leq 1} \mathbb{E}_{\theta}[\| \hat{\theta} - \theta \|_2^2] = (2 + o(1)) \log d, \quad d \to \infty. \]  

(4)
i. (Bayesian lower bound) Consider the prior $\pi$ under which $\theta$ is uniformly distributed over $\{\tau e_1, \ldots, \tau e_d\}$, where $e_i$'s denote the standard basis. Let $\tau = \sqrt{(2 - \epsilon) \log d}$. Show that for any $\epsilon > 0$, the Bayes risk is given by

$$\inf_{\theta} E_{\theta \sim \pi} [\|\hat{\theta} - \theta\|_2^2] = \tau^2(1 + o(1)), \quad d \to \infty.$$ (Hint: use any method you prefer, e.g., mutual information, compute the Bayes risk by evaluating the conditional mean and conditional variance, etc.)

ii. Demonstrate an estimator $\hat{\theta}$ that achieves the RHS of (4) asymptotically. (Hint: one idea is to use hard thresholding $\hat{\theta}_i = X_i 1 \{|X_i| \geq \tau\}$ with appropriately chosen $\tau$.)

(b) To show the lower bound part of (3), prove the generic result

$$R^*(k, d) \geq kR^* \left(1, \frac{d}{k}\right)$$

and apply (3). (Hint: consider a prior of $d/k$ blocks each of which is 1-sparse.)

(c) Demonstrate an estimator $\hat{\theta}$ that achieves the RHS of (3) asymptotically.

Hint: For this problem, the following normal tail bound may be useful:

$$\int_a^\infty \varphi(x) dx \leq \frac{\varphi(a)}{a}.$$  

3. (Non-Gaussian location models) Consider the following location model with iid observations $X_1 = \theta + Z_i$, where $Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} \mu$ for some distribution $\mu$ on $\mathbb{R}$. We know if $\mu$ is standard normal, the minimax quadratic risk is given by $\frac{d}{n}$. To what extent does this rely on Gaussianity?

(a) Assume that $\mu$ satisfies the following conditions: for some universal constant $C$,

- (Quadratic growth of KL) Denote by $T_{\theta}\mu$ the translation of $\mu$ by $\theta$, i.e., the law of $Z + \theta$ if $Z \sim \mu$. Then the KL divergence satisfies

$$D(T_{\theta}\mu \| \mu) \leq C\|\theta\|_2^2, \quad \forall \theta \in \mathbb{R}^d \quad \text{(5)}$$

- $E_{Z \sim \mu} ||Z||_2^2 \leq Cd.$

Show that

$$\inf_{\hat{\theta}} \sup_{\theta \in \mathbb{R}^d} E[\|\hat{\theta} - \theta\|_2^2] \asymp \frac{d}{n}. \quad \text{(6)}$$

(Hint: Fano’s method is useful here for lower bound.)

(b) Show that a sufficient condition for (5) is that $\mu$ has a density $p$ such that the Hessian of $\log \frac{1}{p}$ satisfies $\nabla^2 \log \frac{1}{p} \preceq \frac{C}{2} I_d$.

(c) Here is a counterexample to (6) when (5) fails: Consider $d = 1$ and $\mu = \text{Uniform}(0, 1)$. Explain why in this case (5) does not hold. Show that

$$\inf_{\hat{\theta}} \sup_{\theta \in \mathbb{R}^d} E[\|\hat{\theta} - \theta\|_2^2] \asymp \frac{1}{n^2}. \quad \text{(7)}$$

(Hint: Le Cam’s method is useful here for lower bound; for upper bound, consider the maximum likelihood.)
4. (Shannon lower bound) Let $\| \cdot \|$ be an arbitrary norm on $\mathbb{R}^d$ and $r > 0$. Let $X$ be a $\mathbb{R}^d$-valued random vector with a probability density function $p_X$. Let
\[
F(\epsilon) \triangleq \inf_{P_{\hat{X} \mid X \mid \|X - \hat{X}\| \leq \epsilon}} I(X; \hat{X})
\]
We prove the Shannon lower bound:
\[
F(\epsilon) \geq h(X) + \frac{d}{r} \log \frac{d}{r \epsilon e} - \log \left( \Gamma \left( \frac{d}{r} + 1 \right) V_d \right),
\]
where $h(X) = \int_{\mathbb{R}^d} p_X(x) \frac{1}{p_X(x)} dx$ is the differential entropy of $X$ and $V_d = \text{vol}\{x \in \mathbb{R}^d : \|x\| \leq 1\}$ is the volume of the unit $\| \cdot \|$-ball.

(a) Show that
\[
F(\epsilon) \geq h(X) - G(\epsilon), \quad G(\epsilon) \triangleq \sup_{P_W : \|W\|^r \leq \epsilon} h(W).
\]
(b) Show that
\[
G(\epsilon) = G(1) + \frac{d}{r} \log \epsilon.
\]
(c) Show that $0 < V_d < \infty$.
(d) Show that for any $s > 0$,
\[
Z(s) \triangleq \int_{\mathbb{R}^d} \exp(-s\|w\|^r) dw = \Gamma \left( \frac{d}{r} + 1 \right) V_d s^{-\frac{d}{r}}.
\]
(Hint: $\| \cdot \|$ may not be smooth so you cannot just take derivatives. Use $\int_{\mathbb{R}^d} \exp(-s\|w\|^r) dw = \int_{\mathbb{R}^d} \int_{|w|^r} \infty s \exp(-sx) dx dw$ and apply Fubini’s theorem.)
(e) Consider the probability density function
\[
q_s(w) = \frac{1}{Z(s)} \exp(-s\|w\|^r).
\]
Show that for any feasible $W$ such that $E[\|W\|^r] \leq 1$,
\[
h(W) = -D(p\|q_s) + E \left[ \log \frac{1}{q_s(W)} \right] \leq \log Z(s) + s.
\]
(f) Optimize over $s > 0$ and conclude that
\[
G(1) = -\frac{d}{r} \log \frac{d}{r e} + \log \left( \Gamma \left( \frac{d}{r} + 1 \right) V_d \right).
\]
(g) Assembling all pieces to conclude (8).
(h) (Application of SLB to minimax risk) Consider the $d$-dimensional Gaussian location model:
\[
X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\theta, I_d) \text{ where } \theta \in \Theta \subset \mathbb{R}^d \text{ and } \Theta \text{ has nonempty interior. Apply the mutual information method to prove the following minimax lower bound for quadratic loss:}
\]
\[
R^*_n(\Theta) \triangleq \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}_\theta \|\hat{\theta} - \theta\|^2 \geq \left( \frac{d}{n} \wedge \text{rad}(\Theta)^2 \right) \left( \frac{\text{vol}(\Theta)}{\text{vol}(B_2)} \right)^\frac{2}{d},
\]
where $\text{rad}(\Theta)$ denotes the radius of $\Theta$ with respect to the $\ell_2$-norm and $B_2$ denotes an $\ell_2$-ball of the same radius as $\Theta$. (Hint: choose the uniform prior on $\Theta$.)