Stats 300A HW3 Solutions

Song Mei

October 18, 2018

Problem 1

(a)

Define a projector operator to be the following:

$$\mathrm{Proj}(\boldsymbol{\theta}) = \begin{cases} \boldsymbol{\theta}, & \text{if } \boldsymbol{\theta} \in \Theta^\varepsilon, \\ \arg\min_{\boldsymbol{\theta}' \in \Theta} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2^2, & \text{if } \boldsymbol{\theta} \notin \Theta^\varepsilon. \end{cases}$$

Since Θ is a convex compact set, the minimizer $\arg\min_{\boldsymbol{\theta}' \in \Theta} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2^2$ is unique, so that Proj operator is well defined.

Given an estimator $\hat{\boldsymbol{\theta}} : \mathcal{X} \to \mathbb{R}^d$ such that $\mathsf{P}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) \notin \Theta^\varepsilon) > \delta$, we take $\tilde{\boldsymbol{\theta}} = \mathrm{Proj}(\hat{\boldsymbol{\theta}})$. Then for any $\boldsymbol{\theta} \in \Theta$, we have

$$\|\tilde{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}\|_2^2 \le \|\hat{\boldsymbol{\theta}}(\boldsymbol{x}) - \boldsymbol{\theta}\|_2^2 - \eta \mathbf{1}\{\hat{\boldsymbol{\theta}}(\boldsymbol{x}) \notin \Theta^\varepsilon\},$$

where

$$\eta = \min_{\boldsymbol{\theta} \in \Theta, \boldsymbol{\theta}' \in \partial\Theta^\varepsilon} \|\boldsymbol{\theta}' - \boldsymbol{\theta}\|_2^2 - \|\mathrm{Proj}(\boldsymbol{\theta}') - \boldsymbol{\theta}\|_2^2.$$

Since Θ is a convex compact set, and $\partial\Theta^\varepsilon$ is a compact set, we have $\eta > 0$.

As a result, we have for any $\boldsymbol{\theta} \in \Theta$,

$$\mathsf{E}_{\boldsymbol{\theta}}[\|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\|_2^2] \le \mathsf{E}_{\boldsymbol{\theta}}[\|\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\|_2^2] - \eta \mathsf{P}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}(\boldsymbol{X}) \notin \Theta^\varepsilon) \le \mathsf{E}_{\boldsymbol{\theta}}[\|\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\|_2^2] - \eta\delta,$$

and

$$\sup_{\boldsymbol{\theta} \in \Theta} \mathsf{E}_{\boldsymbol{\theta}}[\|\tilde{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\|_2^2] \le \sup_{\boldsymbol{\theta} \in \Theta} \mathsf{E}_{\boldsymbol{\theta}}[\|\hat{\boldsymbol{\theta}}(\boldsymbol{X}) - \boldsymbol{\theta}\|_2^2] - \eta\delta.$$

That means $\tilde{\boldsymbol{\theta}}$ has strictly better worst risk than $\hat{\boldsymbol{\theta}}$, so that $\hat{\boldsymbol{\theta}}$ is not minimax optimal on Θ.

(b)

First we consider the case when $\boldsymbol{M} \neq \mathbf{0}$. Since Θ is a compact convex set, we take R large enough so that $\Theta^\varepsilon \subseteq \mathsf{B}(\mathbf{0}, R)$ for some small $\varepsilon > 0$. The estimator $\hat{\boldsymbol{\theta}}(\boldsymbol{y}) = \boldsymbol{M}\boldsymbol{y} + \boldsymbol{\theta}_0 \stackrel{d}{=} \boldsymbol{M}\boldsymbol{D}\boldsymbol{\theta} + \boldsymbol{\theta}_0 + \sigma\boldsymbol{M}\boldsymbol{g}$, where $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. Note $\sigma\boldsymbol{M}\boldsymbol{g}$ is not identically $\mathbf{0}$ when $\boldsymbol{M} \neq \mathbf{0}$, and Θ is a compact set, we have

$$\inf_{\boldsymbol{\theta} \in \Theta} \mathsf{P}_{\boldsymbol{\theta}}(\|\boldsymbol{M}\boldsymbol{D}\boldsymbol{y} + \boldsymbol{\theta}_0\|_2 \ge R) \equiv \delta > 0.$$

By problem (a), we conclude that $\hat{\boldsymbol{\theta}}$ cannot be minimax optimal on Θ.

Remark 1. *To show $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ is not minimax optimal, we need to make the additional assumption that $\boldsymbol{D} \in \mathbb{R}^{n \times d}$ has full column rank, otherwise this conclusion doesn't hold. In the following, we prove this conclusion under this additional assumption.*

Then we consider the case when $\boldsymbol{M} = \mathbf{0}$. That means, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$. If $\boldsymbol{\theta}_0 \notin \Theta$, it is obvious $\hat{\boldsymbol{\theta}}$ is not minimax optimal on Θ. Hence we consider the case when $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0 \in \Theta$.

We claim that the $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ cannot be the Bayes estimator for any prior except the prior $\delta(\boldsymbol{\theta}_0)$. Suppose this claim holds, the Bayes risk $R_B(\hat{\boldsymbol{\theta}}, \delta(\theta_0)) = 0$. Since Θ contains at least two points, it is easy to see that the minimax risk should be large than 0, hence $\delta(\theta_0)$ is not the least favorable prior. By minimax theorem, the minimax estimator should be the Bayes estimator for least favorable prior. Therefore, $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ cannot be the minimax estimator.

Now suffice to show the claim above. Suppose Q is a prior probability distribution on Θ and $Q(\Theta \setminus \{\boldsymbol{\theta}_0\}) > 0$, then the Bayes estimator under prior Q and square loss should be the posterior expectation $\hat{\boldsymbol{\theta}}_Q(\boldsymbol{x}) = \mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}]$. We would like to show $\mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}] \not\equiv \boldsymbol{\theta}_0$. The intuition why $\mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}] \not\equiv \boldsymbol{\theta}_0$ can be explained by the following: when $\|\boldsymbol{x}\|_2 \to \infty$, the posterior expectation $\mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}]$ should be at the boundary of the support of Q. In the following we show the above intuition rigorously.

By the fact that $Q(\Theta \setminus \{\boldsymbol{\theta}_0\}) > 0$, there exists a neighborhood $\mathsf{B}(\boldsymbol{\theta}_\star, \delta)$ such that $Q(\mathsf{B}(\boldsymbol{\theta}_\star, \delta)) \equiv \eta > 0$ and $\|\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0\|_2 \geq 2\delta$. Now we take $\boldsymbol{x}_k = \boldsymbol{D}[\boldsymbol{\theta}_0 + k(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0)]$, then we have (denoting $\varphi_n(\boldsymbol{x}) = (1/(2\pi)^{n/2}) \exp\{-\|\boldsymbol{x}\|_2^2/2\}$ to be the standard Gaussian density function on $\mathbb{R}^n$)

$$\langle \mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}_k] - \boldsymbol{\theta}_0, \boldsymbol{\theta}_\star - \boldsymbol{\theta}_0 \rangle = \frac{\int_\Theta \langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, \boldsymbol{\theta}_\star - \boldsymbol{\theta}_0 \rangle \varphi_n(\boldsymbol{D}(\boldsymbol{\theta}_0 - \boldsymbol{\theta} + k(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0))/\sigma) Q(\mathrm{d}\boldsymbol{\theta})}{\int_\Theta \varphi_n(\boldsymbol{D}(\boldsymbol{\theta}_0 - \boldsymbol{\theta} + k(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0))/\sigma) Q(\mathrm{d}\boldsymbol{\theta})}.$$

The integration in the numerator above can be decomposed into the integration in $\mathsf{B}(\boldsymbol{\theta}_\star, \delta)$ and the integration outside $\mathsf{B}(\boldsymbol{\theta}_\star, \delta)$,

$$\begin{aligned}
&\int_\Theta \langle \boldsymbol{\theta} - \boldsymbol{\theta}_0, \boldsymbol{\theta}_\star - \boldsymbol{\theta}_0 \rangle \varphi_n(\boldsymbol{D}(\boldsymbol{\theta}_0 - \boldsymbol{\theta} + k(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0))/\sigma) Q(\mathrm{d}\boldsymbol{\theta}) \\
\geq& \|\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0\|_2 (\|\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0\| - \delta) \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\|\boldsymbol{D}[k(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0) - \boldsymbol{u}]\|_2^2/(2\sigma^2)\}\eta \\
&- \|\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0\|_2 \mathrm{Diam}(\Theta) \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{-\|k\boldsymbol{D}(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0)\|_2^2/(2\sigma^2)\}(1 - \eta),
\end{aligned}$$

where $\mathrm{Diam}(\Theta)$ gives the diameter of Θ, and $\boldsymbol{u} = [(\|\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0\|_2 - \delta)/\|\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0\|_2](\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0)$. Note (we already assumed $\boldsymbol{D}$ has full column rank)

$$\lim_{k \to \infty} \frac{\exp\{-\|\boldsymbol{D}[k(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0) - \boldsymbol{u}]\|_2^2/(2\sigma^2)\}}{\exp\{-\|k\boldsymbol{D}(\boldsymbol{\theta}_\star - \boldsymbol{\theta}_0)\|_2^2/(2\sigma^2)\}} = \infty,$$

hence for large k, we have

$$\langle \mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}_k] - \boldsymbol{\theta}_0, \boldsymbol{\theta}_\star - \boldsymbol{\theta}_0 \rangle > 0.$$

That means, we have $\mathsf{E}_Q[\boldsymbol{\theta}|\boldsymbol{x}_k] \neq \boldsymbol{\theta}_0$ for large k. This proves the claim.

(c)

Let $\Theta = \{-1, 1\}$, $\mathsf{P}_1 = \mathsf{P}_0 = \delta(0)$ (no matter what θ is, the data X is deterministically 0). Hence we only need to consider the estimator that is a constant mapping (Rao-Blackwell theorem tells us that we don't need to consider randomized estimator). The risk function for any constant estimator is $R(\hat{\theta} = a; \Theta) = \sup\{(1-a)^2, (-1-a)^2\}$. Minimizing this over a, the minimax estimator is $\hat{\theta} = 0$. For this estimator, for $\varepsilon < 1/4$, $\mathsf{P}_0(\hat{\theta} \notin \{-1, 1\}^\varepsilon) = \mathsf{P}_1(\hat{\theta} \notin \{-1, 1\}^\varepsilon) = 1$.

(d)

Consider the estimator $\tilde{\theta} = \mathrm{Proj}(\hat{\theta})$, where Proj operator enjoy the same definition of Problem (a), then we have

$$L(\tilde{\theta}(x), \theta) \leq L(\hat{\theta}(x), \theta) - \eta \mathbf{1}\{\hat{\theta}(x) \notin \Theta^\varepsilon\},$$

where

$$\eta = \min_{\theta \in \Theta, \theta' \in \partial\Theta^\varepsilon} L(\theta', \theta) - L(\mathrm{Proj}(\theta'), \theta).$$

Since L is strictly decreasing for $a < \theta$ and strictly increasing for $a > \theta$, and Θ and $\partial\Theta^\varepsilon$ are compact sets, we have $\eta > 0$.

As a result, we have for any $\theta \in \Theta$,

$$R(\tilde{\theta}, \theta) = \mathsf{E}_\theta[L(\tilde{\theta}(X), \theta)] \leq \mathsf{E}_\theta[L(\hat{\theta}(X), \theta)] - \eta \mathsf{P}_\theta(\hat{\theta}(X) \notin \Theta^\varepsilon) \leq R(\hat{\theta}, \theta) - \eta\delta.$$

Since $R(\hat{\theta}, \theta)$ is continuous in θ, $R(\hat{\theta}, \theta)$ can attain the maximum, and we have

$$\sup_{\theta \in \Theta} R(\tilde{\theta}, \theta) \leq \sup_{\theta \in \Theta} R(\hat{\theta}, \theta) - \eta\delta.$$

That means $\hat{\theta}$ is not minimax optimal on Θ.

Problem 2

(a)

Let $\theta_1 = 1/2 - 1/(2\sqrt{2})$, $\theta_2 = 1/2 + 1/(2\sqrt{2})$, and let $q = 1/2$. Under the square loss, the Bayes optimal estimator for Q is given by the conditional expectation

$$\begin{aligned}
\hat{\theta}_B(x) &= \mathsf{E}[\theta | X = x] \\
&= \begin{cases} \frac{\theta_1^2 + \theta_2^2}{\theta_1 + \theta_2} & \text{if } x = 1 \\ \frac{\theta_1(1-\theta_1) + \theta_2(1-\theta_2)}{1-\theta_1+1-\theta_2} & \text{if } x = 0 \end{cases} \\
&= \begin{cases} \frac{3}{4} & \text{if } x = 1 \\ \frac{1}{4} & \text{if } x = 0 \end{cases} \\
&= \frac{x}{2} + \frac{1}{4}.
\end{aligned} \tag{1}$$

The above implies that $\hat{\theta}_B(Q) = \hat{\theta}_{MM}$.

(b)

As suggested in the hint, there exists an integer m, such that choosing $q_i \geq 0$ for $i = 0, 1, \cdots, m$ such that (here Q is the measure induced by a Beta$(\sqrt{n}/2, \sqrt{n}/2)$ random variable)

$$\sum_{i=0}^{m} q_i \left(\frac{i}{m}\right)^k = \int \theta^k Q(\mathrm{d}\theta) \quad \text{for all } k = 0, 1, \cdots n+1. \tag{2}$$

Then the above implies that, for any polynomial p of degree at most $n+1$, we have

$$\sum_{i=0}^{m} q_i p\left(\frac{i}{m}\right) = \int p(\theta) Q(\mathrm{d}\theta). \tag{3}$$

Consider the prior distribution:

$$Q_1 = \sum_{i=0}^{n+1} q_i \delta\left(\frac{i}{m}\right) \tag{4}$$

The Bayes optimal estimator is given by the conditional expectation

$$\begin{aligned}
\hat{\theta}_{Q_1}(X) &= \mathsf{E}_{Q_2}[\theta | X] \\
&= \frac{\sum_{i=0}^{n+1} q_i (i/m)^{X+1} (1 - i/m)^{n-X}}{\sum_{i=0}^{n+1} q_i (i/m)^{X} (1 - i/m)^{n-X}}.
\end{aligned} \tag{5}$$

On the other hand, the Bayes estimator with respect to Beta$(\sqrt{n}/2, \sqrt{n}/2)$ is given by

$$\begin{aligned}
\hat{\theta}_{MM}(X) &= \frac{\sqrt{n}}{1+\sqrt{n}} \cdot \frac{X}{n} + \frac{1}{1+\sqrt{n}} \cdot \frac{1}{2} \\
&= \mathsf{E}_Q[\theta | X] \\
&= \frac{\int \theta^{X+1} (1-\theta)^{n-X} Q(\mathrm{d}\theta)}{\int \theta^{X} (1-\theta)^{n-X} Q(\mathrm{d}\theta)}.
\end{aligned} \tag{6}$$

Let $p_1(t; X) = t^{X+1}(1-t)^{n-X}$, $p_2(t; X) = t^X (1-t)^{n-X}$, then it clear that both p_1 and p_2 as a function of t are polynomial of degree at most $n+1$. Hence by (3) we have

$$\hat{\theta}_{Q_1}(X) = \frac{\sum_{i=0}^{m} p_1\left(\frac{i}{m}, X\right) q_i}{\sum_{i=0}^{m} p_2\left(\frac{i}{m}, X\right) q_i} = \frac{\int p_1(\theta, X) Q(\mathrm{d}\theta)}{\int p_2(\theta, X) Q(\mathrm{d}\theta)} = \hat{\theta}_{MM}(X). \tag{7}$$

Therefore,

$$\hat{\theta}_{Q_1}(X) = \hat{\theta}_{MM}(X) = \frac{\sqrt{n}}{1+\sqrt{n}} \cdot \frac{X}{n} + \frac{1}{1+\sqrt{n}} \cdot \frac{1}{2}. \tag{8}$$

Problem 3

(a)

Since L is upper bounded by L_0, $R(A, \boldsymbol{\theta})$ is also bounded from above by L_0 for all $A \in \mathcal{A}$ and $\boldsymbol{\theta} \in \Theta$. Given Q, for any statistical procedure A, we have

$$\begin{aligned} R(A, Q) = \int_{\mathbb{R}^d} R(A, \boldsymbol{\theta}) Q(\mathrm{d}\boldsymbol{\theta}) &= \int_{\Theta} R(A, \boldsymbol{\theta}) Q(\mathrm{d}\boldsymbol{\theta}) + \int_{\Theta^c} R(A, \boldsymbol{\theta}) Q(\mathrm{d}\boldsymbol{\theta}) \\ &\leq \sup_{\boldsymbol{\theta} \in \Theta} R(A, \boldsymbol{\theta}) + L_0 Q(\Theta^c). \end{aligned} \tag{9}$$

Hence

$$R_B(Q) - L_0 Q(\Theta^c) \leq R(A, Q) - L_0 Q(\Theta^c) \leq \sup_{\boldsymbol{\theta} \in \Theta} R(A, \boldsymbol{\theta}). \tag{10}$$

Since the above is true for all A, taking the infimum over $A \in \mathcal{A}$ gives

$$R_M(\Theta) \geq R_B(Q) - L_0 Q(\Theta^c). \tag{11}$$

(b)

Let $\hat{\boldsymbol{\theta}}$ be any estimator, and let $\tilde{\boldsymbol{\theta}}$ be the projection of $\hat{\boldsymbol{\theta}}$ onto $\mathsf{B}^d(\mathbf{0}, M\sqrt{k})$. That is

$$\tilde{\boldsymbol{\theta}} = \min\left\{ \frac{M\sqrt{k}}{\|\hat{\boldsymbol{\theta}}\|_2}, 1 \right\} \hat{\boldsymbol{\theta}}. \tag{12}$$

Then it is clear that $L(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) \leq L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$ with probability 1 for all $\boldsymbol{\theta} \in \Theta(d, k, M) \subset \mathsf{B}^d(\mathbf{0}, M\sqrt{k})$. Since $\tilde{\boldsymbol{\theta}} \in \mathsf{B}^d(\mathbf{0}, M\sqrt{k})$, it is sufficient to only consider estimators taking values in $\mathsf{B}^d(\mathbf{0}, M\sqrt{k})$. In this case, since both $\tilde{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}$ are in a ball with radius $M\sqrt{k}$, there distance square is upper bounded by the diameter square of the ball. That is, for all $\boldsymbol{\theta} \in \Theta$ and $\tilde{\boldsymbol{\theta}}$ in the above form, we have

$$L(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}) \leq 4M^2 k. \tag{13}$$

Therefore it is also sufficient to replace the square loss by $\tilde{L}(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \min\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2, 4M^2 k\}$.

(c)

Let $G = \Pi_d \times \Sigma_d$ be a group, where Π_d is the permutation group on $\{1, \ldots, d\}$, and $\Sigma_d = \{+1, -1\}^d$ is the sign changing group. For any $g = [\pi, \boldsymbol{\sigma}] \in G$ (π is a permutation, where $\{\pi(1), \ldots, \pi(d)\} = \{1, \ldots, d\}$ as a set; $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_d]^\mathsf{T} \in \{+1, -1\}^d$), the action of φ_g on $\boldsymbol{x} = (x_1, \ldots, x_d)^\mathsf{T} \in \mathbb{R}^d$ gives $\varphi_g(\boldsymbol{x}) = (\sigma_1 x_{\pi(1)}, \ldots \sigma_d x_{\pi(d)})^\mathsf{T}$. We would like to show our statistical model is invariant under this group. First we have $L(a, \boldsymbol{\theta}) = \|a - \boldsymbol{\theta}\|_2^2 = \|\varphi_g(a) - \varphi_g(\boldsymbol{\theta})\|_2^2 = L(\varphi_g(a), \varphi_g(\boldsymbol{\theta}))$. Next we have $\mathsf{P}_{g(\boldsymbol{\theta})}(\boldsymbol{X} \in S) = \mathsf{P}_{\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)}(\varphi_g(\boldsymbol{\theta}) + \boldsymbol{Z} \in S) = \mathsf{P}_{\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)}(\varphi_g(\boldsymbol{\theta}) + \varphi_g(\boldsymbol{Z}) \in S) = \mathsf{P}_{\boldsymbol{Z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_d)}(\varphi_g(\boldsymbol{\theta} + \boldsymbol{Z}) \in S) = \mathsf{P}_{\boldsymbol{\theta}}(\varphi_g(\boldsymbol{X}) \in S) = (\varphi_g)_{\#} \mathsf{P}_{\boldsymbol{\theta}}(\boldsymbol{X} \in S)$. Hence our model is invariant under this group. Since minimax theorem holds for this model, there exists a least favorable prior. According to invariant least favorable prior theorem, there exists a least favorable prior that is invariant under the group action. This invariant least favorable prior can only be written in the form $Q = \sum_{\ell=0}^{k} p_\ell Q_\ell$.

(d)

By part (b) we know that $R_M(d, k; M) = \tilde{R}_M(d, k; M)$, and we can replace the loss L by $\tilde{L}$, which is bounded from above by $4M^2 k$. By part (a) we have

$$R_M(d, k; M) = \tilde{R}_M(d, k; M) \geq \tilde{R}_B(Q_{M,\epsilon}) - 4M^2 k Q_{M,\epsilon}(\Theta^c). \tag{14}$$

Let $\boldsymbol{X} \in \mathbb{R}^d$ be a random variable whose induced measure is $Q_{M,\epsilon}$, then it is clear that $Q_{M,\epsilon}(\Theta^c)$ is equal to $\mathsf{P}(\|\boldsymbol{X}\|_0 > k)$. Since the coordinates of $\boldsymbol{X}$ are independent and $\mathbf{1}(X_i \neq 0)$ has Bernoulli(ϵ) distribution, $\|\boldsymbol{X}\|_0$ has Binomial(d, ϵ) distribution. Therefore, (14) becomes

$$R_M(d, k; M) \geq \tilde{R}_B(Q_{M,\epsilon}) - 4M^2 k \mathsf{P}(\mathrm{Binom}(d, \epsilon) > k). \tag{15}$$

(e)

Note $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \sim Q_{M,\varepsilon} = q_{M,\varepsilon}^{\otimes d}$, and $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\theta}, \sigma^2 \mathbf{I}_d)$. We have (X_i, θ_i) for $i \in [d]$ are mutually independent. Hence the Bayes estimator which is the posterior mean gives

$$(\hat{\boldsymbol{\theta}}_B(\boldsymbol{x}))_j = \mathsf{E}[\theta_j | \boldsymbol{X} = \boldsymbol{x}] = \mathsf{E}[\theta_j | X_j = x_j].$$

Hence

$$\begin{aligned}
R_B(Q_{M,\epsilon}) &= \mathsf{E}_{Q_{M,\epsilon}}[\|\hat{\boldsymbol{\theta}}_B - \boldsymbol{\theta}\|_2^2] \\
&= \sum_{j \in [d]} \mathsf{E}_{Q_{M,\epsilon}}[((\hat{\boldsymbol{\theta}}_B)_j - \theta_j)^2] \\
&= \sum_{j \in [d]} \mathsf{E}_{Q_{M,\epsilon}}[(\mathsf{E}(\theta_j | X_j) - \theta_j)^2] \\
&= \sum_{j \in [d]} \mathsf{E}_{q_{M,\epsilon}}[(\mathsf{E}(\theta_j | X_j) - \theta_j)^2] \\
&= d R_B(q_{M,\epsilon}).
\end{aligned} \tag{16}$$

Since we have

$$P(\text{Binom}(d, \epsilon) > k) \leq e^{-k\eta^2/4}, \tag{17}$$

which implies that $kP(\text{Binom}(d, \epsilon) > k) = o_\eta(k)$, using (15), (16) and (7) in the question gives

$$R_M(d, k; M) \geq d R_B(q_{M,\epsilon}) - (M^2 + 1) o_\eta(k) - 4M^2 o_\eta(k). \tag{18}$$

Since a constant times $o_\eta(k)$ is still $o_\eta(k)$, the $-4M^2 o_\eta(k)$ above can be merged with the first $M^2 o_\eta(k)$, so it can be simplifies to

$$R_M(d, k; M) \geq d R_B(q_{M,\epsilon}) - (M^2 + 1) o_\eta(k). \tag{19}$$