Stats 300A

Suyash Gupta Homework 4 solutions

#1: Small noise limit in the bounded normal mean model

Solution.

(a) The estimator X has a constant risk function σ^2. So,

$$R(\Theta; \sigma^2) \leq \sigma^2 . \tag{1}$$

(b) We use choice 1. Under the given prior, the posterior distribution $\theta|x$ has a truncated Normal Distribution. Precisely speaking,

$$\theta|X \sim Y| - 1 < Y < 1$$

where $Y \sim N(X, \sigma^2)$

From wiki, the mean of the posterior distribution is given by

$$X + \sigma \frac{\varphi(\frac{-1-X}{\sigma}) - \varphi(\frac{1-X}{\sigma})}{\Phi(\frac{1-X}{\sigma}) - \Phi(\frac{-1-X}{\sigma})}$$

which is the required Bayes estimator.

Now,

$$\frac{1}{\sigma^2} R(Q_\sigma) = \mathsf{E}\left(\frac{X-\theta}{\sigma} + \frac{\varphi(\frac{-1-X}{\sigma}) - \varphi(\frac{1-X}{\sigma})}{\Phi(\frac{1-X}{\sigma}) - \Phi(\frac{-1-X}{\sigma})}\right)^2$$

Take $Z = \frac{X-\theta}{\sigma}$.

$$\frac{1}{\sigma^2} R(Q_\sigma) = \mathsf{E}\left(Z + \frac{\varphi(\frac{-1-\theta}{\sigma} - Z) - \varphi(\frac{1-\theta}{\sigma} - Z)}{\Phi(\frac{1-\theta}{\sigma} - Z) - \Phi(\frac{-1-\theta}{\sigma} - Z)}\right)^2$$

$$\begin{aligned}
&\liminf_{\sigma\to 0} \frac{1}{\sigma^2} R(Q_\sigma) \\
&\qquad = \liminf_{\sigma\to 0} \mathsf{E}\left(Z + \frac{\varphi(\frac{-1-\theta}{\sigma} - Z) - \varphi(\frac{1-\theta}{\sigma} - Z)}{\Phi(\frac{1-\theta}{\sigma} - Z) - \Phi(\frac{-1-\theta}{\sigma} - Z)}\right)^2 \\
&\qquad \geq \mathsf{E} \liminf_{\sigma\to 0} \left(Z + \frac{\varphi(\frac{-1-\theta}{\sigma} - Z) - \varphi(\frac{1-\theta}{\sigma} - Z)}{\Phi(\frac{1-\theta}{\sigma} - Z) - \Phi(\frac{-1-\theta}{\sigma} - Z)}\right)^2 \\
&\qquad = EZ^2 = 1
\end{aligned}$$

where the last inequality is by Fatou's lemma and the last equality is by continuity of Normal density and distribution function.

(c) From part (a),

$$\limsup_{\sigma\to 0} \frac{1}{\sigma^2} R(Q_\sigma) \leq 1$$

Hence, combining with b). we have the result.

□

#2: A modified James–Stein Estimator

Solution.

(a) Here $g(x) = \hat{\theta}(x) - x = -(x - \bar{x}\mathbf{1})h(||x - \bar{x}||^2)$. So,

$$\begin{aligned} R(\hat{\theta}, \theta) &= \mathsf{E}(||\hat{\theta} - \theta||^2) \\ &= d + \mathsf{E}(||g(x)||^2) + 2\mathsf{E}(div(g(x))) \end{aligned}$$

Let $f(x) = ||x - \bar{x}\mathbf{1}||^2$, so $div(g(x) = -(d-1)h(f) - 2fh'(f)$.

So,

$$R(\hat{\theta}, \theta) = d + \mathsf{E}[fh(f)^2 - 2(d-1)h(f) + 2fh'(f)]$$

(b) Plugging in the required values of h and h', we get,

$R(\hat{\theta}, \theta) = d + (C^2 - 2dC + 6C)\mathsf{E}(\frac{1}{f})$

(c) We want $C^2 - 2dC + 6C = C(C - 2d + 6) < 0$. If $d \geq 3$, then taking $C \in (0.2d - 6)$ gives the strict inequality.

(d) Consider the model where θ_i are i.i.d. samples from $N(\mu, \sigma^2)$, and $X_i|\theta_i \sim N(\theta_i, 1)$. Following the empirical Bayes procedure, we estimate μ, σ^2 using moment method. The marginal distribution of X_i is $N(\mu, 1 + \sigma^2)$, so $\hat{\mu} = \bar{X}$ and $\hat{\sigma^2} = ||X - \bar{X}||^2/C - 1$.

For squared error loss, Bayes estimate is

$$\hat{\theta_{Bayes}}(x) = \frac{\mu}{1 + \sigma^2}\mathbf{1} + \frac{\sigma^2}{1 + \sigma^2}X$$

Plugging in the estimates for $\hat{\mu}$ and $\hat{\sigma^2}$, we have that $\hat{\theta_{Bayes}}(x)$ is actually the modified James Stein estimator.

□

#3: A regression problem with random designs

Solution. **Thanks to Zi Yang Kang for solution to Problem 3**

(a) Let $\hat{\boldsymbol{\theta}}$ be a minimax estimator. Suppose that for given data $(\boldsymbol{y}, \mathbf{X})$, $\mathrm{dist}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}); \Theta) > 1$, where $\mathrm{dist}(x; S) = \inf\{\|x - s\|_2 : s \in S\}$. Then consider the procedure $\widetilde{\boldsymbol{\theta}}$ defined by

$$\widetilde{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}) = \begin{cases} \hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}) & \text{for } \mathrm{dist}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}); \Theta) \leq 1, \\ \mathbf{0} & \text{for } \mathrm{dist}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}); \Theta) > 1. \end{cases}$$

Observe that

$$\|\widetilde{\boldsymbol{\theta}}(y, \boldsymbol{x}) - \boldsymbol{\theta}\|_2^2 = \begin{cases} \|\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}) - \boldsymbol{\theta}\|_2^2 & \text{for } \mathrm{dist}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}); \Theta) \leq 1, \\ 1 & \text{for } \mathrm{dist}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}); \Theta) > 1. \end{cases}$$

Consequently, $R(\widetilde{\boldsymbol{\theta}}; \boldsymbol{\theta}) \leq R(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$ for every $\boldsymbol{\theta} \in \Theta$; hence, if $\hat{\boldsymbol{\theta}}$ is minimax optimal, then $\widetilde{\boldsymbol{\theta}}$ is minimax also. Therefore, to construct a minimax estimator $\hat{\boldsymbol{\theta}}$, it suffices to consider $\hat{\boldsymbol{\theta}}$ such that $\mathrm{dist}(\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}); \Theta) \leq 1$. Equivalently, it suffices to consider $\hat{\boldsymbol{\theta}}$ such that $\mathrm{im}(\hat{\boldsymbol{\theta}}) \subseteq B^d(2)$, the d-dimensional closed ball with radius 2.

(b) By part (a), we may assume without loss of generality that the action space $\mathcal{A} = B^d(2)$. Since $\mathcal{A}$ and Θ are compact and L is continuous, the minimax theorem applies; hence a least favorable prior $\overline{\mathsf{Q}}_* \in \mathscr{M}_1(\Theta)$ exists.

Consider the d-dimensional orthogonal group, $O(d)$, endowed with its rotation action on Θ and $\mathcal{A}$ (*i.e.*, $\varphi_Q : \boldsymbol{v} \mapsto Q\boldsymbol{v}$ for $Q \in O(d)$ and $\boldsymbol{v} \in \Theta, \mathcal{A}$) and its anti-rotation action on $\mathcal{X}$ (*i.e.*, $\varphi_Q : \boldsymbol{x} \mapsto Q^\top \boldsymbol{x}$ for $Q \in O(d)$ for $\boldsymbol{x} \in \mathcal{X}$). We claim that the model is invariant under $O(d)$. Indeed, observe that

$$\begin{aligned} (\varphi_Q)_\# \mathsf{P}_{\boldsymbol{\theta}}\{(\boldsymbol{y}, \mathbf{X}) \in S\} &= (\varphi_Q)_\# \mathsf{P}_{\boldsymbol{\theta}}\{((y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)) \in S\} \\ &= \mathsf{P}_{\boldsymbol{\theta}}\{((y_1, Q^\top \boldsymbol{x}_1), (y_2, Q^\top \boldsymbol{x}_2), \ldots, (y_n, Q^\top \boldsymbol{x}_n)) \in S\} \\ &= \mathsf{P}_{Q\boldsymbol{\theta}}\{((y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)) \in S\} = \mathsf{P}_{\varphi_Q(\boldsymbol{\theta})}\{(\boldsymbol{y}, \mathbf{X}) \in S\}. \end{aligned}$$

Moreover, for all $\boldsymbol{a} \in \mathcal{A}$, $\boldsymbol{\theta} \in \Theta$ and $Q \in O(d)$,

$$\|\varphi_Q(\boldsymbol{a}) - \varphi_Q(\boldsymbol{\theta})\|_2^2 = \|Q\boldsymbol{a} - Q\boldsymbol{\theta}\|_2^2 = \|\boldsymbol{a} - \boldsymbol{\theta}\|_2^2.$$

Therefore the model is invariant under $O(d)$. Since $O(d)$ is compact and there exists a least favorable prior $\overline{\mathsf{Q}}_*$, hence there exists a least favorable prior Q_* that is invariant under the action $\varphi_Q : \boldsymbol{\theta} \mapsto Q\boldsymbol{\theta}$ on Θ. This means that $(\varphi_Q)_\# \mathsf{Q}_* = \mathsf{Q}_*$ for every $Q \in O(d)$. The only $\mathsf{Q}_* \in \mathscr{M}_1(\Theta)$ satisfying this is the uniform prior.

We thus conclude that a least favorable prior Q_* is the uniform prior on Θ, such that $\mathsf{Q}_*(\{\boldsymbol{\theta}_1\}) = \mathsf{Q}_*(\{\boldsymbol{\theta}_2\})$ for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$.

(c) As justified in part (b), the minimax theorem applies; therefore, the minimax estimator is the Bayes estimator with respect to the least favorable prior $\mathsf{Q}_* = \mathsf{Q}_{\mathsf{unif}}$.

Under square loss, the Bayes (and minimax) estimator $\hat{\boldsymbol{\theta}}_M$ is given by

$$\hat{\boldsymbol{\theta}}_M(\boldsymbol{y}, \mathbf{X}) = \mathbb{E}\{\boldsymbol{\theta} \,|\, \boldsymbol{y}, \mathbf{X}\}.$$

Here, the expectation is taken conditional on the data $(\boldsymbol{y}, \mathbf{X})$, with respect to the (unconditional) measure $\mathsf{Q}_* = \mathsf{Q}_{\mathsf{unif}}$ on Θ.

(d) We compute that

$$\mathsf{E}_{\boldsymbol{\theta}} \,||\, \hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}) - \boldsymbol{\theta} \,||_2^2 \; = 1 - \frac{2}{C(n)} \sum_{i=1}^{n} \mathsf{E}_{\boldsymbol{\theta}} \,\langle y_i \boldsymbol{x}_i, \boldsymbol{\theta} \rangle + \frac{1}{[C(n)]^2} \,\mathsf{E}_{\boldsymbol{\theta}} \left\{ \sum_{i=1}^{n} ||\, y_i \boldsymbol{x}_i \,||_2^2 \right\}.$$

Independence of w_k and $\boldsymbol{x}_k$ implies that, for any distinct $i, j \in \{1, 2, \ldots, d\}$,

$$\begin{aligned} \mathsf{E}_{\boldsymbol{\theta}} \,\langle y_i \boldsymbol{x}_i, y_j \boldsymbol{x}_j \rangle &= \mathsf{E}_{\boldsymbol{\theta}} \left\langle \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{\theta} + w_i \boldsymbol{x}_i, \boldsymbol{x}_j \boldsymbol{x}_j^\top \boldsymbol{\theta} + w_j \boldsymbol{x}_j \right\rangle \\ &= \mathsf{E}_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{x}_j \boldsymbol{x}_j^\top \boldsymbol{\theta} \right\} \\ &= \mathsf{E}_{\boldsymbol{\theta}} \left\{ \theta_1^2 x_{i,1}^2 x_{j,1}^2 + \theta_2^2 x_{i,2}^2 x_{j,2}^2 + \cdots + \theta_d^2 x_{i,d}^2 x_{j,d}^2 \right\} = 1. \end{aligned}$$

Similarly,

$$\begin{aligned} \mathsf{E}_{\boldsymbol{\theta}} \,||\, y_i \boldsymbol{x}_i \,||_2^2 &= \mathsf{E}_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{\theta} + w_i^2 \boldsymbol{x}_i^\top \boldsymbol{x}_i \right\} \\ &= \mathsf{E}_{\boldsymbol{\theta}} \left\{ \left(\theta_1^2 x_{i,1}^2 + \theta_2^2 x_{i,2}^2 + \cdots + \theta_d^2 x_{i,d}^2 \right) \left(x_{i,1}^2 + x_{i,2}^2 + \cdots + x_{i,d}^2 \right) \right\} + \sigma^2 \,\mathsf{E}_{\boldsymbol{\theta}} \,||\, \boldsymbol{x}_i \,||_2^2 \\ &= d + 2 + \sigma^2 d. \end{aligned}$$

On the other hand,

$$\mathsf{E}_{\boldsymbol{\theta}} \langle y_i \boldsymbol{x}_i, \boldsymbol{\theta} \rangle = \mathsf{E}_{\boldsymbol{\theta}} \left\langle \boldsymbol{x}_i \boldsymbol{x}_i^\top \boldsymbol{\theta} + w_i \boldsymbol{x}_i, \boldsymbol{\theta} \right\rangle = \mathsf{E}_{\boldsymbol{\theta}} \left\{ \theta_1^2 x_{i,1}^2 + \theta_2^2 x_{i,2}^2 + \cdots + \theta_d^2 x_{i,d}^2 \right\} = 1.$$

Combining our above computations yields

$$\mathsf{E}_{\boldsymbol{\theta}} \, || \, \hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}) - \boldsymbol{\theta} \, ||_2^2 = 1 - \frac{2n}{C(n)} + \frac{n\,(d + 2 + \sigma^2 d) + n\,(n-1)}{[C(n)]^2}$$

Simplifying:

$$R(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}) = 1 - n \cdot \frac{2\,C(n) - n - d - \sigma^2 d - 1}{[C(n)]^2}.$$

We wish to minimize $R(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta})$. We adopt the following approach. Consider the maximization problem:

$$\max_{z \in \mathbb{R}} \frac{2z - n - d - \sigma^2 d - 1}{z^2}.$$

Using elementary calculus, we find that the solution z_* must satisfy

$$2z_*^2 - 2z_* \left(2z_* - n - d - \sigma^2 d - 1\right) = 0 \implies 2z_* \left(n + d + \sigma^2 + 1 - z_*\right) = 0.$$

Observe that

$$\frac{2z - n - d - \sigma^2 d - 1}{z^2} \to -\infty \quad \text{as } z \to 0.$$

Thus the solution to the maximization problem is $z_* = n + d + \sigma^2 d + 1$, for which

$$\frac{2z_* - n - d - \sigma^2 d - 1}{z_*^2} = \frac{1}{n + d + \sigma^2 d + 1}.$$

Therefore we optimally set $C(n) = n + d + \sigma^2 d + 1$:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{y}, \mathbf{X}) = \frac{1}{n + d + \sigma^2 d + 1} \sum_{i=1}^{n} y_i \boldsymbol{x}_i \implies R(\hat{\boldsymbol{\theta}}; \boldsymbol{\theta}) = \frac{d + \sigma^2 d + 1}{n + d + \sigma^2 d + 1}.$$

We conclude that an upper bound for the minimax risk is

$$R_M(\Theta) \leq \frac{d + \sigma^2 d + 1}{n + d + \sigma^2 d + 1}.$$

□