Statistical Theory for High and Infinite Dimensions

April 19, 2017

3.18: Rate of convergence of lasso

Normal means model,

$$Y_i = \theta_i + \varepsilon_i,\ i = 1, \ldots, n,\ \varepsilon_i \overset{\text{iid}}{\sim} N(0,1)$$

Oracle: θ is s-sparse. Lasso estimate given by

$$\hat{\theta}_j = \begin{cases} Y_i - \lambda/2 & Y_i > \lambda/2 \\ 0 & |Y_i| \leq \lambda/2 \\ Y_i + \lambda/2 & Y_i < -\lambda/2 \end{cases}$$

Which λ should we choose? Define 'best' λ to be the one that minimizes the risk

$$\mathbb{E}_\theta \left\| \hat{\theta} - \theta \right\|_2^2 = \text{bias}(\hat{\theta})^2 + \text{Var}(\hat{\theta})$$

Let $S = \{j : \theta_j \neq 0\}$ (unknown!), with $|S| = s$ (known).

Bias

Suppose $j \notin S \Rightarrow \theta_j = 0$. By symmetry of Y_j,

$$\text{bias}(\hat{\theta}_j) = \mathbb{E}_\theta \hat{\theta}_j = 0$$

Suppose $j \in S$. In the worst case, the lasso estimate shifts the unbiased estimates Y_i by $\lambda/2$. Hence,

$$\text{bias}(\hat{\theta}_j)^2 \leq \lambda^2/4$$

Thus, the squared bias term in the MSE is $\leq s\lambda^2/4$.

Variance Part 1

Start again with $j \notin S$. For the variance, write

$$\hat{\theta}_j = (Y_j - \lambda/2) \cdot 1_{Y_j > \lambda/2} + (Y_j + \lambda/2) \cdot 1_{Y_j < -\lambda/2}$$

Then, (by unbiased property)

$$\text{Var}_\theta \hat{\theta}_j = \mathbb{E}_\theta \hat{\theta}_j^2 = \mathbb{E}_\theta \left[(Y_j - \lambda/2)^2 \cdot 1_{\{Y_j > \lambda/2\}} \right] + \mathbb{E}_\theta \left[(Y_j + \lambda/2)^2 \cdot 1_{\{Y_j < -\lambda/2\}} \right]$$

Both terms are equal by symmetry. Writing out the first element on RHS:

$$\mathbb{E}_\theta \left[Y_j^2 1_{\{Y_j > \lambda/2\}} \right] - \lambda \mathbb{E}_\theta \left[Y_j 1_{\{Y_j > \lambda/2\}} \right] + \frac{\lambda^2}{4} \mathbb{E}_\theta \left[1_{\{Y_j > \lambda/2\}} \right]$$

Exercise 3.16*: for Z standard normal, $p > 0$, there exists a C_p such that for all $a \geq 1$,

$$\mathbb{E} Z^p 1_{Z > a} \leq C_p a^p e^{-\frac{1}{2}a^2} \lesssim a^p e^{-\frac{1}{2}a^2}$$

This gives, for $\lambda > 2$,

$$\mathbb{E}_\theta \left[Y_j^2 1_{\{Y_j > \lambda/2\}} \right] \lesssim \lambda^2 e^{-\frac{\lambda^2}{8}}$$

We also have

$$\lambda \cdot \mathbb{E}_\theta \left[Y_j 1_{\{Y_j > \lambda/2\}} \right] \overset{*}{\lesssim} \lambda^2 e^{-\frac{\lambda^2}{8}}$$

and

$$\frac{\lambda^2}{4} \cdot \mathbb{E}_\theta \left[1_{\{Y_j > \lambda/2\}} \right] \overset{**}{\lesssim} \lambda^2 e^{-\frac{\lambda^2}{8}}$$

So the first part of the variance is $\lesssim (n-s)\lambda^2 e^{-\frac{\lambda^2}{8}}$.

Variance part 2

Now let $j \in S$. Rewrite

$$\hat{\theta}_j = Y_j - Y_j \cdot 1_{|Y_j| \leq \lambda/2} - \frac{\lambda}{2} \cdot 1_{Y_j > \lambda/2} + \frac{\lambda}{2} \cdot 1_{Y_j < -\lambda/2}$$

Then $\text{Var}\left[\hat{\theta}_j\right] \leq \mathbb{E}_\theta \left[\left(\hat{\theta}_j - \theta_j \right)^2 \right]$

$$= \mathbb{E}_\theta \left[\left(Y_j - \theta_j - Y_j \cdot 1_{|Y_j| \leq \lambda/2} - \frac{\lambda}{2} \cdot 1_{Y_j > \lambda/2} + \frac{\lambda}{2} \cdot 1_{Y_j < -\lambda/2} \right)^2 \right]$$

$$\leq 4\mathbb{E}_\theta \left[\left(Y_j - \theta_j \right)^2 \right] + 4\mathbb{E}_\theta \left[\left(Y_j \cdot 1_{|Y_j| \leq \lambda/2} \right)^2 \right]$$

$$+ 4\mathbb{E}_\theta \left[\left(\frac{\lambda}{2} \cdot 1_{Y_j > \lambda/2} \right)^2 \right] + 4\mathbb{E}_\theta \left[\left(\frac{\lambda}{2} \cdot 1_{Y_j < -\lambda/2} \right)^2 \right]$$

$$\mathbb{E}_\theta\left[\left(Y_j - \theta_j\right)^2\right] = 1$$

$$\mathbb{E}_\theta\left[\left(Y_j \cdot 1_{|Y_j|\leq\lambda/2}\right)^2\right] \leq \frac{\lambda^2}{4}$$

$$\mathbb{E}_\theta\left[\left(\frac{\lambda}{2} \cdot 1_{Y_j>\lambda/2}\right)^2\right] \leq \frac{\lambda^2}{4}$$

$$\mathbb{E}_\theta\left[\left(\frac{\lambda}{2} \cdot 1_{Y_j<-\lambda/2}\right)^2\right] \leq \frac{\lambda^2}{4}$$

Hence, second part of the variance is $\leq 2 \cdot s(4 + 3\lambda^2)$

End

Together, we have

$$\begin{aligned} MSE(\hat{\theta}_\lambda) &\lesssim s\lambda^2/4 + 2 \cdot s(4 + 3\lambda^2) + (n - s)\lambda^2 e^{-\frac{\lambda^2}{8}} & (1) \\ &= C_1 s + C_2 s\lambda^2 + (n - s)\lambda^2 e^{-\frac{\lambda^2}{8}} & (2) \end{aligned}$$

For $(\lambda^*)^2 = 8\log(n/s)$, risk is bounded by a multiple of $s\log(n/s)$, provided $s = o(n)$.

Exercise 3.17*

Exercise 3.17: for Z standard normal, $p > 0$, there exists a C_p such that for all $a \geq 1$,

$$\mathbb{E}Z^p 1_{Z>a} \leq C_p a^p e^{-\frac{1}{2}a^2}$$

Proof Let Z be standard normal, $p \leq 1$ and $a > 1$.

$$\begin{aligned} \mathbb{E}Z^p 1_{Z>a} &= \frac{1}{\sqrt{2\pi}} \int_a^\infty z^p e^{-\frac{z^2}{2}} dz \\ &\propto \int_a^\infty z^p e^{-\frac{z^2}{2}} dz \\ &\leq \int_a^\infty \left(\frac{z}{a}\right)^{1-p} z^p e^{-\frac{z^2}{2}} dz \\ &= a^{p-1} \int_a^\infty z e^{-\frac{z^2}{2}} dz \\ &= a^{p-1} e^{-\frac{a^2}{2}} \end{aligned}$$

Thus

$$\mathbb{E}Z^p 1_{Z>a} \lesssim a^{p-1} e^{-\frac{a^2}{2}}$$

. Let $p > 1$ and $a \geq 1$. Then, by integration by parts,

$$\begin{aligned}
\mathbb{E}Z^p 1_{Z>a} &\propto \int_a^\infty z^p e^{-\frac{z^2}{2}} dz \\
&= \int_a^\infty z^{p-1} \cdot z \, e^{-\frac{z^2}{2}} dz \\
&= \left[-z^{p-1} e^{-\frac{z^2}{2}} \right]_a^\infty + (p-1) \int_a^\infty z^{p-2} e^{-\frac{z^2}{2}} dz \\
&= a^{p-1} e^{-\frac{1}{2}a^2} + (p-1) \int_a^\infty z^{p-2} e^{-\frac{z^2}{2}} dz
\end{aligned}$$

. By iterating, we obtain

$$\begin{aligned}
\mathbb{E}Z^p 1_{Z>a} &\propto a^{p-1} e^{-\frac{1}{2}a^2} + (p-1) a^{p-3} e^{-\frac{1}{2}a^2} \\
&+ (p-1)(p-3) a^{p-5} e^{-\frac{1}{2}a^2} + \ldots + \text{remainder} \\
&\leq C_p a^{p-1} e^{-\frac{1}{2}a^2}
\end{aligned}$$