Risk Inflation relative to Bayes Oracle

Charles Zheng and Yuval Benjamini

October 22, 2015

These are preliminary notes.

1 Ridge regression

Suppose $\beta \sim N(0, \frac{\sigma^2\alpha^2}{p} I)$, $X \sim N^n(0, I_p)$ and $y \sim N(X\beta, \sigma^2 I_n)$.

Define the risk of ridge regression as

$$R(\lambda) = \mathbf{E}||y^* - (x^*)^T \hat{\beta}_\lambda||^2$$

where

$$\hat{\beta}_\lambda = (X^T X + n\lambda)^{-1} X^T y$$

and where $x^* \sim N(0, I_p)$, $y^* \sim ((x^*)^T \beta, \sigma^2)$.

Knowing α^2, one should set $\lambda = \lambda^*$, by

$$\lambda^* = \frac{\gamma}{\alpha^2} = \frac{(p/n)}{\alpha^2}.$$

However, for α^2 unknown, we propose the following. Choose a constant c, and set $\lambda = \lambda_c$, where

$$\lambda_c = c\frac{||y||^2}{n}$$

Let $R^* = \mathbf{E}R(\lambda^*)$ and $R_c = \mathbf{E}R(\lambda_c) = \mathbf{E}R(c||y||^2/n)$.

Claim: fixing c,

$$\sup_{\alpha^2 \geq 0} \frac{R_c}{R^*} < \infty$$

where $\gamma = p/n$.

Implication: simply setting $\lambda = c||y||^2/n$, one can achieve a risk that is at worst a bounded multiple of the risk of the optimal rule for choosing λ.

For now we prove a weaker, asymptotic version of the claim where

$$\max\{\lim_{\alpha^2\to\infty}\frac{R_c}{R^*}, \frac{R_c}{R^*}\bigg|_{\alpha^2=0}\} < \infty$$

in the limit where α^2, σ^2 are fixed while $n \to \infty$, $p \to \infty$, and $p/n \to \gamma$.

1.1 Preliminaries

Define

$$Q = (\gamma - \lambda - 1)^2 + 4\gamma\lambda$$

In the limit, we have

$$R(\lambda) = \frac{1}{2\gamma}\left[\alpha^2(\gamma - 1 - \sqrt{Q} - \lambda\left(\frac{1+\lambda+\gamma}{\sqrt{Q}}\right)) + \gamma\left(1 + \frac{1+\lambda+\gamma}{\sqrt{Q}}\right)\right]$$

and in particular that

$$R(\lambda^*) = R(\gamma\alpha^{-2}) = \frac{1}{2}\left[1 + \frac{\gamma-1}{\gamma}\alpha^2 + \sqrt{(1+\frac{\gamma-1}{\gamma}\alpha^2)^2 + 4\alpha^2}\right]$$

2 Covariance estimation

$$S \sim W_n(\frac{1}{n}\Sigma), D = \mathrm{diag}(S), \hat{R} = D^{-1/2}SD^{-1/2}$$

$$S_\lambda = \lambda D + (1-\lambda)S$$

Which λ minimizes

$$\mathbf{E}\mathrm{tr}[S_\lambda^{-1}\Sigma] + \log\det(S_\lambda)$$

We have

$$\log\det(\lambda D + (1-\lambda)S) = \log\det D + \log\det(\lambda I + (1-\lambda)\hat{R}) = \log\det D + \sum_{i=1}^{p}\log(\lambda + (1-\lambda)r_i)$$

where r_i are the eigenvalues of $\hat{R}$.

Meanwhile

$$\begin{aligned}\mathbf{E}\mathrm{tr}[S_\lambda^{-1}\Sigma] &= \mathbf{E}\mathrm{tr}[(\lambda D + (1-\lambda)S)^{-1}\Sigma] \\ &= \frac{1}{1-\lambda}\mathbf{E}\mathrm{tr}[(\frac{\lambda}{1-\lambda}I + \hat{R})^{-1}D^{-1/2}\Sigma D^{-1/2}]\end{aligned}$$

Take $n, p \to \infty$. Then $D^{-1/2}\Sigma D^{-1/2} \to R$, the true correlation. From now we can just assume $\Sigma = R$, (ie unit marginal variances), it doesn't matter in the limit. Then we get

$$\mathbf{E}\mathrm{tr}[S_\lambda^{-1}\Sigma] = \frac{1}{1-\lambda}\mathbf{E}\mathrm{tr}[(\frac{\lambda}{1-\lambda}I + S)^{-1}\Sigma]$$

We know how to evaluate the term inside, i.e. $\mathbf{E}\mathrm{tr}[(\frac{\lambda}{1-\lambda}I + S)^{-1}\Sigma]$ based on random matrix theory.