Risk Inflation relative to Bayes Oracle

Charles Zheng and Yuval Benjamini

October 24, 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 $\lambda = g(X)$ for some function of X. We will consider

$$g(X) = \text{const}$$

Let $R^* = \mathbf{E}R(\lambda^*)$ and $R_\lambda = \mathbf{E}R(\lambda)$.

Claim: fixing λ,

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

where $\gamma = p/n$.

Implication: simply using a fixed choice of λ, 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

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

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

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.

3 Minimum of non-central chi-squared

Problem: investigate sequences $d \to \infty$, $L \to \infty$, and $\lambda \to \infty$ so that the distribution of

$$M = \min_{i=1}^{L} ||Z_i||^2$$

converges in probability to a constant, where $Z \sim N(\sqrt{\lambda}e_1, I_d)$.

If M does converge to to a constant, the p-th quantile of M should also converge to that constant, for any $p \in (0,1)$. But the pth quantile of M is also the $1 - \sqrt[L]{(1-p)}$th quantile of $X = ||Z||^2$. Hence we would like to compute

$$q_p(d, \lambda, L) = \{x : \Pr[X < x] = 1 - \sqrt[L]{1-p}\}$$

First, we can approximate

$$\Pr[X < x] \approx \mathrm{Vol}(B_{\sqrt{x}})\phi_d(\lambda e_1) = \frac{2^{1-(d/2)}x^{d/2}}{\Gamma(d/2)d}e^{-\lambda/2}$$

where ϕ is the central multivariate normal density.

Lemma. *For any d, λ we have*

$$\lim_{x\to 0}\frac{\frac{2^{1-(d/2)}x^{d/2}}{\Gamma(d/2)d}e^{-\lambda/2}}{\Pr[||Z||^2 < x]} = 1$$

Proof. Write

$$\Pr[||Z||^2 < x] = \int_{B_{\sqrt{x}}} (2\pi)^{-d/2} e^{-||z-\sqrt{\lambda}e_1||^2/2} dz$$

We can approximate the integrand by $(2\pi)^{-d/2}e^{-\lambda/2}$, since

$$\sup_{||z||^2<x} \frac{|e^{-||z-\sqrt{\lambda}e_1||^2/2} - e^{-\lambda/2}|}{e^{-\lambda/2}} = |e^{\sqrt{x}/2} - 1| = O(\sqrt{x})$$

Hence,

$$\begin{aligned}\Pr[||Z||^2 < x] &= \int_{B_{\sqrt{x}}} (2\pi)^{-d/2} e^{-||z-\sqrt{\lambda}e_1||^2/2} dz \\ &= (1 + O(\sqrt{x})) \int_{B_{\sqrt{x}}} (2\pi)^{-d/2} e^{-\lambda/2} dz = (1 + O(\sqrt{x}))\mathrm{Vol}(B_{\sqrt{x}})\phi_d(\lambda e_1)\end{aligned}$$

which completes the proof.

Secondly, we can approximate

$$1 - \sqrt[L]{1-p} \approx -\frac{\log(1-p)}{L}$$

Hence, the pth quantile of M is x which approximately satisfies

$$-\frac{\log(1-p)}{L} = \frac{2^{1-(d/2)}x^{d/2}}{\Gamma(d/2)d} e^{-\lambda/2}$$

Meaning

$$\begin{aligned} q_p(d, \lambda, L) &\approx \left(-\frac{\log(1-p)}{L} \frac{\Gamma(d/2)d}{2^{1-(d/2)}} e^{\lambda/2} \right)^{2/d} \\ &\approx \left(-\frac{\log(1-p)}{L} \frac{\sqrt{\pi d}(d/(2e))^{d/2} d}{2^{1-(d/2)}} e^{\lambda/2} \right)^{2/d} \\ &= \left(-\frac{\log(1-p)}{L} \right)^{2/d} \frac{d^{1+(3/d)} \sqrt[d]{\pi}}{2^{2/d}} e^{(\lambda/d)-1} \end{aligned}$$