STAT 666 J&R Theorem 1

Maggie Buffum

May 28, 2019

Theorem 1. Consider the ridge traces $\{\hat{\beta}_R^i(k); k \geq 0\}$, and their variances $\{Var(\hat{\beta}_R^i(k)); k \geq 0\}$, and the derivatives $\{dVar(\hat{\beta}_R^i(k))/dk; k \geq 0\}$, for $1 \leq i \leq p$. Then,

(i) Variances are given by $\{Var(\hat{\beta}_R^i(k)) = \sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^2)\mathbf{a}_i;\ 1 \leq i \leq p\}$, each monotone decreasing with increasing k.

Consider the ridge regression form $\{(\mathbf{Z}'\mathbf{Z} + k\mathbf{I}_P)\omega = \mathbf{Z}'\mathbf{Y}_0\}$. Written in correlation form (that is, centered and scaled to the unit), the ridge solutions are found by solving $\{(\mathbf{D}_\delta^2 + k\mathbf{I}_p)\theta = \mathbf{D}_\delta \mathbf{W}\}$, which yields $\hat{\theta}_R(k) = \mathbf{D}(\delta_i(\delta_i^2 + k))\mathbf{W}$ and $\hat{\omega}_R(k) = \mathbf{Q}\hat{\theta}_R(k)$, where $\mathbf{D}(\delta_i(\delta_i^2 + k))$ is a diagonal matrix with $\delta_i(\delta_i^2 + k)$ along the diagonals and $\mathbf{P}\mathbf{D}_\delta\mathbf{Q}'$ diagonalizes $\mathbf{Z}$.

Now we have that $\hat{\beta}_R(k) = \mathbf{A}\hat{\theta}_R(k)$. It's easy to see that

$$Var(\hat{\theta}_R(k)) = \sigma^2 \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^2) \implies Var(\hat{\beta}_R(k)) = \sigma^2 \mathbf{A}\mathbf{D}(\delta_i^2/(\delta_i^2 + k)^2)\mathbf{A}'.$$

The monotonicity of $Var(\hat{\beta}_R^i(k))$ is proven in (ii) by showing that the derivative of $Var(\hat{\beta}_R^i(k))$ is negative with respect to k.

(ii) Rates of change in $\{Var(\hat{\beta}_R^i(k)));\ 1 \leq i \leq p\}$ are given by

$$\{dVar(\hat{\beta}_R^i))/dk = -2\sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^3)\mathbf{a}_i < 0;\ 1 \leq i \leq p\}$$

independently of $\mathbf{Y}$.

Using the chain rule, it's clear that

$$\begin{aligned}
\{dVar(\hat{\beta}_R^i(k))/dk &= (d/dk)\sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^2)\mathbf{a}_i\} \\
&= \{(d/dk)\sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2(\delta_i^2 + k)^{-2})\mathbf{a}_i\} \\
&= \{-2\sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2(\delta_i^2 + k)^{-3})\mathbf{a}_i\} \\
&= \{-2\sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^3)\mathbf{a}_i\}
\end{aligned}$$

(iii) The negative functions $\{dVar(\hat{\beta}_R^i(k))/dk;\ 1 \leq i \leq p\}$ are monotone increasing as k increases for $k \geq 0$, their values progressing from large to small in magnitude.

Taking the second derivative of $Var(\hat{\beta}_R^i(k))$ with respect to k, we have (similar to [ii]),

$$\begin{aligned}
d^2Var(\hat{\beta}_R^i(k))/dk^2 &= (d^2/dk^2) - \sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^3)\mathbf{a}_i \\
&= 6\sigma^2 \mathbf{a}_i' \mathbf{D}(\delta_i^2/(\delta_i^2 + k)^4)\mathbf{a}_i.
\end{aligned}$$

Therefore, $\{dVar(\hat{\beta}_R^i(k))/dk\}$ are monotone increasing functions of k.