STA 721 HW 10

1. Verify equation (3) starting from (1) for the rotated model for $\mathbf{Y}^*$:

$$\mathbf{Y} = \mathbf{1}\alpha + \mathbf{X}^s\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{1}$$

$$\mathbf{U}^T\mathbf{Y} = \mathbf{U}^T\mathbf{1}\alpha + \mathbf{U}^T\mathbf{X}^s\boldsymbol{\beta} + \mathbf{U}^T\boldsymbol{\epsilon} \tag{2}$$

$$\begin{pmatrix} y_0^* \\ y_1^* \\ \vdots \\ y_p^* \\ y_{p+1}^* \\ \vdots \\ y_{n-1}^* \end{pmatrix} = \begin{bmatrix} \sqrt{n} & 0 & 0 & \dots & 0 \\ 0 & l_1 & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0 & \dots \\ \vdots & 0 & \vdots & \ddots & 0 \\ \vdots & \vdots & \vdots & 0 & l_p \\ \vdots & \vdots & \vdots & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \dots & 0 \end{bmatrix} \begin{pmatrix} \alpha \\ \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix} + \boldsymbol{\epsilon}^* \tag{3}$$

where$\mathbf{X}^s$ has been centered and standardized so that each column has length 1 and mean 0 (see scale in R), $\mathbf{U} = [\mathbf{1}_n/\sqrt{n}\,\mathbf{U}_p\,\mathbf{U}_{n-p-1}]$ is an orthogonal matrix with $\mathbf{X}^s = \mathbf{U}_p L \mathbf{V}^T$ from the singular value decomposition, and $\boldsymbol{\gamma} = \mathbf{V}^T\boldsymbol{\beta}$.

2. Show that OLS estimates $\hat{\alpha} = \bar{y}$ and $\hat{\gamma}_i = y_i^*/l_i$ for $i = 1, \dots, p$.

3. Show that $\mathsf{SSE} = \mathbf{Y}^T\mathbf{U}_{n-p-1}\mathbf{U}_{n-p-1}^T\mathbf{Y} = \mathbf{Y}(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T - \mathbf{U}_p\mathbf{U}_p^T)\mathbf{Y}$. Do you need to find $\mathbf{U}_{n-p-1}^T\mathbf{Y}$ to compute SSE? Can you express SSE as a function of the MLEs of α and $\hat{\gamma}$s (and $\mathbf{Y}^T Y$)?

4. Write the likelihood function as a product of three terms: a part that involves only $\alpha \mid \phi$, $\gamma \mid \phi$ and ϕ with the MLES and SSE using $\mathbf{Y}^*$.

5. Derive the full conditional distributions for α, γ, κ_j and ϕ assuming

$$p(\alpha, \phi) \propto 1/\phi \tag{4}$$

$$\gamma_j \mid \kappa_j, \phi, \alpha \overset{\text{iid}}{\sim} \mathsf{N}(0, \frac{1}{\phi\kappa_j}) \tag{5}$$

$$\kappa_j \overset{\text{iid}}{\sim} G(1/2, 1/2) \tag{6}$$

You should have a name for the distribution and expressions for all hyperparameters, not just an expression for the density. Hint: write down likelihoods and priors, but ignore any terms that do not involve the parameter of interest (they go into constant of proportionality). Simplify until you recognize the distribution.

6. Modify your Gamma prior on κ_i to try to capture the desired features of Goldstein & Smith (1974) (See Christensen Chapter 15) where if

$$\gamma_i^2 < \sigma^2 \left[\frac{2}{\kappa_i} + \frac{1}{l_i^2} \right]$$

the fixed κ_i Generalized Ridge shrinkage estimator (posterior mean) beats OLS:

- If l_i is small almost any κ_i will improve over OLS
- if l_i^2 is large then only very small values of κ_i will give an improvement based on l_i, i.e. if l_i^2 is large $kappa_i$ should be small. Plot your prior densities, with an overlay of the G(1/2, 1/2).

7. Find the updated full conditionals based on your choice above. Do you need to update all of the full conditionals? Explain.

8. Implement your 2 models in R, JAGS or other language (see earlier JAGS code as a starting point) and apply this to the `longley` data. How do your results compare to classical ridge? Include histograms of the posterior distributions of coefficients, plus means and credible intervals, as well as histograms of the κ's with the prior density overlaid. How sensitive are the results to the prior assumptions? How do the estimates of κ_i compare to the best GCV estimate from class?

9. Explain the computational advantage of using the canonical parameterization in MCMC.