Math 583 HW 5 2023 Due Wednesday, Oct. 4. Exam 1 Wed. Sept. 27.

1) Let the full model $Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_{p-1} X_{p-1} + e$. Suppose that a submodel I uses the constant and $k-1$ nontrivial predictors $X_{i1}, ..., X_{i,k-1}$. Let $X_{i,k}, X_{i,k+1}, ..., X_{i,p-1}$ denote the predictors left out of the model. Then the partial F test statistic F_I tests whether submodel I is good or whether at least one of the predictors left out of the model is needed. Let $\boldsymbol{r}$ denote the residuals from the full model, let $\boldsymbol{r}_I$ denote the residuals from the submodel, let $C_p(I)$ denote the C_p criterion for the submodel I, and let n be the sample size. Then it can be shown that

$$\text{corr}(\boldsymbol{r}, \boldsymbol{r}_{\text{I}}) = \sqrt{\frac{\text{n}-\text{p}}{\text{C}_\text{p}(\text{I}) + \text{n} - 2\text{k}}} = \sqrt{\frac{\text{n}-\text{p}}{(\text{p}-\text{k})\text{F}_\text{I} + \text{n} - \text{p}}}.$$

Assume that $-p \leq C_p(I) \leq k$ and $0 \leq F_I \leq 1$. Then what happens to $\text{corr}(\boldsymbol{r}, \boldsymbol{r}_{\text{I}})$ as $n \to \infty$?

2) For ridge regression, let $\boldsymbol{A}_n = (\boldsymbol{X}^T\boldsymbol{X} + \lambda_{1,n}\boldsymbol{I}_p)^{-1}\boldsymbol{X}^T\boldsymbol{X}$ and $\boldsymbol{B}_n = [\boldsymbol{I}_p - \lambda_{1,n}(\boldsymbol{X}^T\boldsymbol{X} + \lambda_{1,n}\boldsymbol{I}_p)^{-1}]$. Show $\boldsymbol{A}_n - \boldsymbol{B}_n = \boldsymbol{0}$.

3) Suppose $n = 12$ and observations are measured on the following people.
1) Ahlam, 2) Mani, 3) Abdul, 4) Kasun, 5) Lakni, 6) Sanjuka, 7) Siraj, 8) Paul, 9) Mina, 10) James, 11) Seth, 12) An.

```
sample(1:12)
 [1] 11 12  5  7  2  9  3  1 10  4  6  8
```

Use the above output to determine which people are in the training set H that uses $n_H = 6$ cases.

4) If $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$, then the OLS estimator minimizes $Q_{OLS}(\boldsymbol{\beta}) = RSS(\boldsymbol{\beta}) = \|\boldsymbol{Y} - \boldsymbol{X\beta}\|_2^2 = (\boldsymbol{Y} - \boldsymbol{X\beta})^T(\boldsymbol{Y} - \boldsymbol{X\beta}) = \boldsymbol{Y}^T\boldsymbol{Y} - 2\boldsymbol{Y}^T\boldsymbol{X\beta} + \boldsymbol{\beta}^T(\boldsymbol{X}^T\boldsymbol{X})\boldsymbol{\beta}$. Using Theorem 3.4 (exam 2 review 68)) with $\boldsymbol{a}^T = -2\boldsymbol{Y}^T\boldsymbol{X}$, $\boldsymbol{A} = \boldsymbol{X}^T\boldsymbol{X}$, and $\boldsymbol{\eta} = \boldsymbol{\beta}$, find the gradient $\triangledown Q(\boldsymbol{\beta})$. (Setting the gradient = equal to zero and solving for $\hat{\boldsymbol{\beta}}$ shows that the OLS estimator satisfies the normal equations $(\boldsymbol{X}^T\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{Y}$.)

5) The function `predsim2` simulates the data splitting prediction region. The arguments are dtype = 1 for $(T, \boldsymbol{C}) = (\overline{\boldsymbol{x}}, \boldsymbol{I}_p)$ and dtype = 2 for $(T, \boldsymbol{C}) = (MED(\boldsymbol{W}), \boldsymbol{I}_p)$, xtype = 1 for $\boldsymbol{x} \sim N_p(\boldsymbol{0}, diag(1, 2, ..., p))$, 2 for lognormal (if $\boldsymbol{w} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$, then $x_i = exp(w_i)$ for $i = 1, ..., p$), and xtype = 3 for $\boldsymbol{x} \sim N_p(\boldsymbol{0}, \boldsymbol{I})$. The function uses $n_V = min(n_V, floor(n/2))$ and $n_H = n - n_V$. The simulation uses 100 runs. Hence 100 data sets are generated with $\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \boldsymbol{x}_f$ iid. The simulation gives cvr=the proportion of the 100 nominal 95% prediction regions, formed from the training data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$, that contain the test data vector $\boldsymbol{x}_f$. The actual coverage up is also given, where $U_V = min(n_V, ceiling[(1-\alpha)(1+n_V)])$ and $up = U_V/(n_V+1)$. with $n_V = 19$, the actual coverage and nominal coverage of 0.95 are the same. There is no reason to believe that the observed coverage cvr differs from the actual coverage up=0.95 if $0.89 \leq cvr \leq 1$.

Copy and paste the R commands for this problem into R. Do not forget the two source commands. For each part, record cvr. For each part, $n = 50$ and xtype = 2 (lognormal data). The computational time increases with p.

a) Here $p = 5$ and dtype =1.
b) Here $p = 5$ and dtype =2.
c) Here $p = 50$ and dtype =1.
d) Here $p = 50$ and dtype =2.
e) Here $p = 500$ and dtype =1.
f) Here $p = 500$ and dtype =2.