MATH 437/537 ASSIGNMENT VI Due March 4, 2020

The dataset `www.mines.edu/~wnavidi/math437537/concrete.csv` contains data from 28 construction jobs involving the construction of concrete silos. Three of the variables describe resource requirements. These are the volume of concrete required in m^3 (y), the number of crew-days of labor (z), or the number of concrete mixer hours (w) needed for a particular job. The table below defines 23 potential independent variables that can be used to predict y, z, or w.

x_1	Number of bins	x_{13}	Breadth to thickness ratio
x_2	Maximum required concrete per hr.	x_{14}	Perimeter of complex
x_3	Height	x_{15}	Mixer capacity
x_4	Sliding Rate of the Slipform (m/day)	x_{16}	Density of stored material
x_5	Number of construction stages	x_{17}	Waste percent in reinforcing steel
x_6	Perimeter of slipform	x_{18}	Waste percent in concrete
x_7	Volume of silo complex	x_{19}	Number of workers in concrete crew
x_8	Surface area of silo walls	x_{20}	Wall thickness (cm)
x_9	Volume of one bin	x_{21}	Number of reinforcing steel crews
x_{10}	Wall-to-floor areas	x_{22}	Number of workers in forms crew
x_{11}	Number of lifting jacks	x_{23}	Length to breadth ratio
x_{12}	Length to thickness ratio		

1. Let $\mathbf{X}$ be the matrix whose columns are the variables x_1–x_{23}. Let $\mathbf{Y}$ be the volume of concrete required (y in the data set). Fit the full model $\mathbf{Y} = \beta_0\mathbf{1}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$ and obtain the least-squares estimate $\hat{\boldsymbol{\beta}}$.

2. Let $\mathbf{W}$ be the standardized version of $\mathbf{X}$. Construct the matrix $\mathbf{Z}$ whose columns are the principal components of $\mathbf{W}$. Fit the model $\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma}+\boldsymbol{\varepsilon}$, and find the least-squares estimator $\hat{\gamma}$.

3. How many principal components are necessary to explain 90% of the variation in $\mathbf{W}$?

4. Let k be the number of principal components needed to explain at least 90% of the variation in $\mathbf{W}$. Find the reduced estimator $\hat{\boldsymbol{\gamma}}_k$.

5. Compute the fitted values $\hat{\mathbf{Y}}$ from the fit of the full model and from the reduced model. Plot the estimates from the full model against those from the fitted model. Are they similar?

6. Use cross-validation to choose the number of principal components. Because there are only 28 observations, use the leave-one-out method, This can be done by specifying `validation="LOO"`. How many principal components does this method suggest to use?

7. Let

$$\mathbf{X} = \begin{bmatrix} 10 & 8 & 8 & 7 & 8 \\ 11 & 6 & 8 & 9 & 5 \\ 7 & 7 & 5 & 4 & 10 \\ 4 & 4 & 0 & 10 & 2 \\ 9 & 12 & 6 & 8 & 6 \\ 8 & 7 & 9 & 9 & 7 \\ 3 & 4 & 7 & 10 & 7 \end{bmatrix} \qquad \mathbf{Y} = \begin{bmatrix} -76 \\ 24 \\ 27 \\ -8 \\ 8 \\ 37 \\ -12 \end{bmatrix}$$

Assume the model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ holds.

(a) Find the value of R^2 when $\mathbf{Y}$ is regressed on $\mathbf{X}$.

(b) How much of the variation in $\mathbf{X}$ is explained by the first four principal components?

(c) Find the value of R^2 when $\mathbf{Y}$ is regressed on the first four principal components of $\mathbf{X}$.

Required for MATH 537, extra credit for MATH 437:

8. Let $\mathbf{X}$ be the matrix in problem 1. Find a vector $\mathbf{Y}$ such that the value of R^2 when $\mathbf{Y}$ is regressed on $\mathbf{X}$ is 1, and the value of R^2 when $\mathbf{Y}$ is regressed on the first 22 principal components of $\mathbf{X}$ is 0.