ACTU PS5841 Data Science in Finance & Insurance - Spring 2022 (Y. Wang)
Assignment - 6
Assigned 2/24/22, Due 3/6/22 (Sun)

Problem 1. LASSO
The date file, *data.csv*, contains observations generated by

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \epsilon_i$$

You decide to investigate three regression models:
a null model $E(y_i|x_i) = \beta_0$ with the least squares loss
a full model $E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}$ with the least squares loss
a LASSO model $E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5}$ with the LASSO loss

You also decide to fit these models using *centered-and-normalized features* with each feature having a zero mean and a unit L2-norm.

Please code to calculate and report the following in your writeup (see the report template):

Model	λ_{best}	$\lvert\hat{\boldsymbol{\beta}}\rvert$	$RMSE_{fit}$	R^2_{fit}	$E(RMSE_{test})$	$E(R^2_{test})$
Null	na	na				
Full	na					
LASSO at λ_{best} by LassoCV						
LASSO at λ_{best} by YourCV w. rescaling						
LASSO at λ_{best} by YourCV w/o rescaling						

Model	β_0	β_1	β_2	β_3	β_4	β_5
Null						
Full						
LASSO at λ_{best} by LassoCV						
LASSO at λ_{best} by YourCV w. rescaling						
LASSO at λ_{best} by YourCV w/o rescaling						

For convenience of grading,
A. Please use a 10-fold CV when estimating $E(RMSE_{test})$ and $E(R^2_{test})$ for the Null model and the Full model.

B. $|\hat{\boldsymbol{\beta}}| = \sum_{j=1}^{5} |\beta_j|$ does NOT include the intercept.

C. λ_{best} by LassoCV is obtained by

```
LassoCV(alphas = lambdas, cv = KFold(n_splits = 10, shuffle = True, random_state = 10))
```

D. λ_{best} by YourCV is obtained by identifying the λ from $\boldsymbol{\lambda} = \text{np.linspace}(0, 0.1, 101)$ that has the smallest estimated test RMSE, using a 10-fold cross validation approach. Please use

```
KFold(n_splits = 10, shuffle = True, random_state = 10)
```

You may find the following resources useful.

```
https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Lasso.html
```

```
https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LassoCV.html
```