10-Assignment BrundageKelley

Kelley Brundage

7/17/2019

The summary results of Assignment #5 Code are below which is needed for the completion of Assignment #10 as it pulls from the ELS Train and ELSE Test datasets. The code below is a compilation of a linear model (mod_5) of reading scores by Socioeconomic Status, coefficient of that lm, RMSE of that model and then a linear model (mod_6) of reading scores as a factor of race and Socioeconomic Status, predictions of the second model and the RMSE of the second model. These are items needed in order to complete the homework assignment.

Question 1:

Create a 10-fold cross validation of your linear model predicting reading scores as a function of at least two covariates. Provide a summary table or graphic of the RMSEs from this cross validation.

Define and Run the first Model which is from the ELS_Train dataset and is looking at reading scores compared to the two covariates of Socioeconomic Status and Race.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.17	0.2982	84.4	0
byses1	5.076	0.1332	38.12	3.469e-291
byrace	0.7714	0.05156	14.96	6.756e-50

Table 2: Fitting linear model: mod1_formula

Observations	Residual Std. Error	R^2	Adjusted R^2
7642	8.44	0.2098	0.2096

ELS_Train converted to tibbles resulting in rmse for model 1:

Table 3: RMSE Apply Model 1

.id	rmse
01	8.725
02	8.017
03	8.029
04	8.377
05	8.394
06	8.618
07	8.518
08	8.551
09	8.519
10	8.644

The RMSE for Model 1 reflects:

1	2	3	4	5	6	7	8	9	10
8.725	8.017	8.029	8.377	8.394	8.618	8.518	8.551	8.519	8.644

The resulting dataset includes the id for the cross validation and the RMSE We can summarize and plot this new data frame to see what our likely range of RMSE happens to be.

Density Plot for RMSE Dataframe
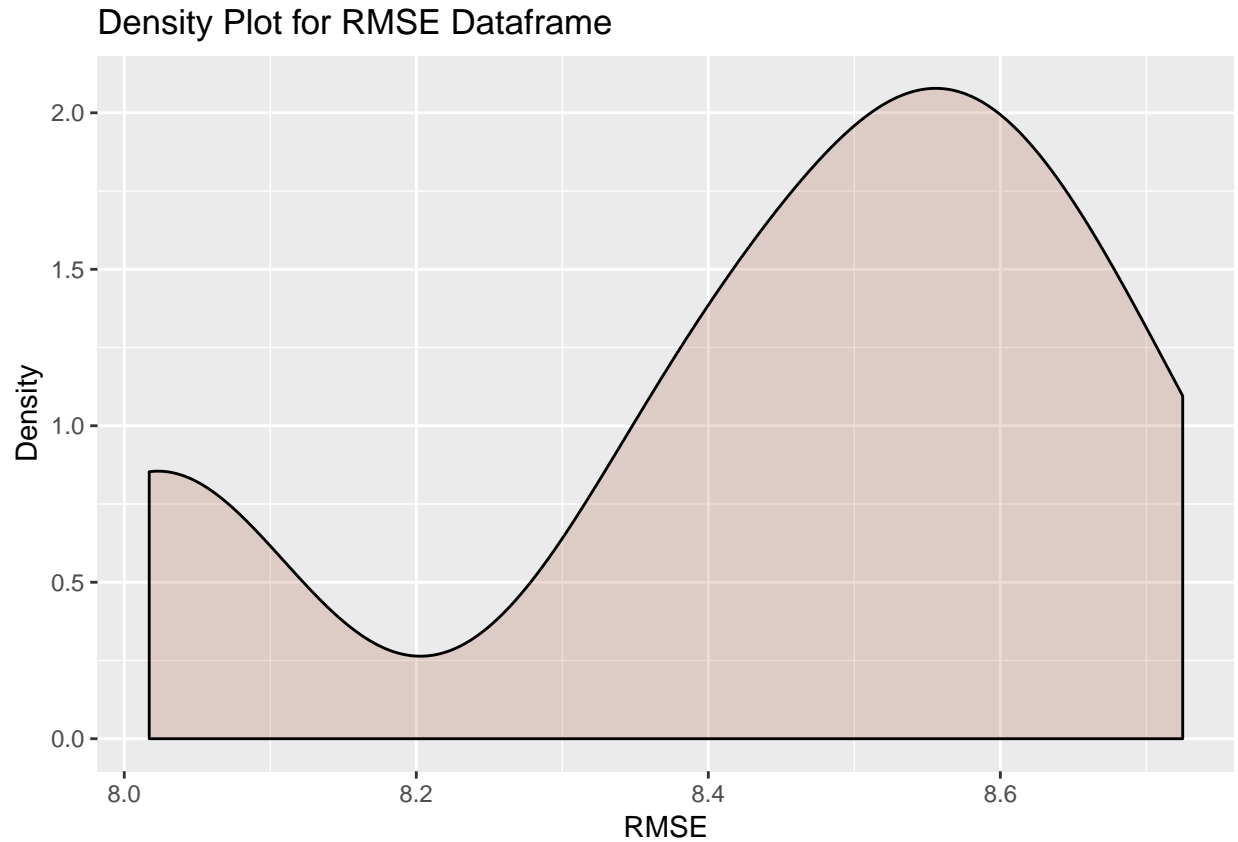
Question 2:

Using a random partition, create 100 separate cross validations of your linear model predicting reading scores as a function of at least two covariates. Provide a summary table or graphic of the RMSEs from this cross validation.

ELS_Train converted to tibbles resulting in rmse for model 1 cross validation:

.id	rmse
Length:100	Min. :8.171
Class :character	1st Qu.:8.417
Mode :character	Median :8.469
NA	Mean :8.469
NA	3rd Qu.:8.544
NA	Max. :8.745

Within the cross validation of the RMSE Model 1 we see:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
8.171	8.417	8.469	8.469	8.544	8.745

Below is a density plot for the RMSE Model 1 with cross validation:

Plot of Model 1 RMSE Cross Validation
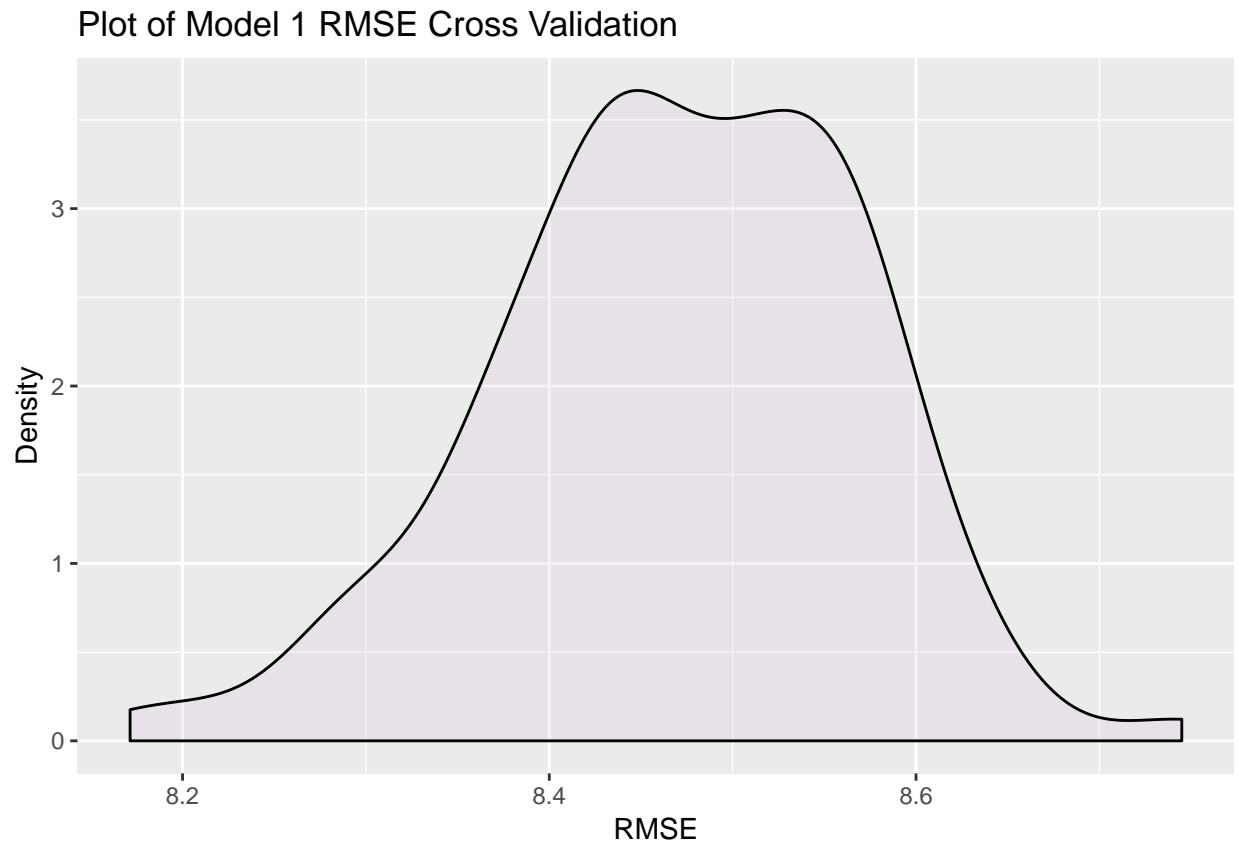
Below is a density plot comparing the two cross validation models:

Comparison of Cross Validation Models