Bike renting
Sankepally vikram reddy
12/7/2018

Contents:

Model results

[1] "ridge"
[1] " model on casual count"
<mark>Ridge Regression</mark>

500 samples
 33 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 449, 449, 451, 449, 451, 450, ...
Resampling results across tuning parameters:

lambda	RMSE	Rsquared	MAE
0e+00	335.0056	0.7110740	243.1953
1e-04	335.0042	0.7110768	243.1931
1e-01	339.0056	0.7074126	247.1517

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was lambda = 1e-04.
[1] "test RMSE of casual count prediction"
[1] 426.2306
[1] "model on registered model"
[1] "registered count"
Ridge Regression

500 samples
 29 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 450, 451, 449, 450, 451, 450, ...
Resampling results across tuning parameters:

lambda	RMSE	Rsquared	MAE
0e+00	480.7583	0.8382233	361.2641
1e-04	480.7457	0.8382312	361.2606
1e-01	488.1834	0.8345103	370.3040

RMSE was used to select the optimal model using the smallest value.

The final value used for the model was lambda = 1e-04.
[1] "test RMSE of registered prediction"
[1] 896.3894

[1] "Test RMSE on Total count(casual +registered)"
[1] 1873.119

[1] "lm"
[1] " model on casual count"
Linear Regression

500 samples
 30 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 450, 448, 450, 451, 451, 449, ...
Resampling results:

  RMSE   Rsquared MAE
  310.6568 0.719439 231.7858

Tuning parameter 'intercept' was held constant at a value of TRUE
[1] "test RMSE of casual count prediction"
[1] 345.5146
[1] "model on registered model"
[1] "registered count"
Linear Regression

500 samples
 29 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 450, 451, 450, 449, 450, 451, ...
Resampling results:

  RMSE   Rsquared   MAE
  481.7545 0.8402131 362.0418

Tuning parameter 'intercept' was held constant at a value of TRUE
[1] "test RMSE of registered prediction"

[1] 896.4213

[1] "lasso"
[1] " model on casual count"
The lasso

500 samples
 30 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 449, 451, 449, 451, 449, 449, ...
Resampling results across tuning parameters:

fraction	RMSE	Rsquared	MAE
0.1	425.0802	0.6214633	318.4660
0.5	319.3486	0.7073917	235.5467
0.9	308.3910	0.7224082	230.7805

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was fraction = 0.9.
[1] "test RMSE of casual count prediction"
[1] 345.5146
[1] "model on registered model"
[1] "registered count"
The lasso

500 samples
 29 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 449, 450, 450, 451, 448, 450, ...
Resampling results across tuning parameters:

fraction	RMSE	Rsquared	MAE
0.1	762.3609	0.6945451	618.3600
0.5	504.8551	0.8290416	387.7460
0.9	480.1704	0.8397033	361.3526

RMSE was used to select the optimal model using the smallest value.

The final value used for the model was fraction = 0.9.
[1] "test RMSE of registered prediction"
[1] 896.4213
[1] "Test RMSE on Total count(casual +registered)"
[1] 1842.415

[1] "glm"
[1] " model on casual count"
Generalized Linear Model

500 samples
 30 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 451, 449, 450, 449, 450, 450, ...
Resampling results:

  RMSE   Rsquared   MAE
  310.1827 0.7214924 232.12

[1] "test RMSE of casual count prediction"
[1] 345.5146
[1] "model on registered model"
[1] "registered count"
Generalized Linear Model

500 samples
 29 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 451, 450, 450, 450, 451, 451, ...
Resampling results:

  RMSE   Rsquared   MAE
  481.5329 0.8393977 362.6352

[1] "test RMSE of registered prediction"
[1] 896.4213
[1] "Test RMSE on Total count(casual +registered)"
[1] 1842.415

[1] "enet"
[1] " model on casual count"
Elasticnet

500 samples
 30 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 450, 451, 450, 452, 449, 449, ...
Resampling results across tuning parameters:

lambda	fraction	RMSE	Rsquared	MAE
0e+00	0.050	469.2667	0.5926228	357.1032
0e+00	0.525	320.8617	0.7112272	234.7054
0e+00	1.000	310.0510	0.7228402	232.2465
1e-04	0.050	538.4552	0.5456557	410.5538
1e-04	0.525	336.9844	0.6840356	242.2628
1e-04	1.000	310.0489	0.7228449	232.2454
1e-01	0.050	537.9420	0.5491185	410.2240
1e-01	0.525	336.2158	0.6824469	242.6476
1e-01	1.000	312.5700	0.7219067	236.3410

RMSE was used to select the optimal model using the smallest value.
The final values used for the model were fraction = 1 and lambda = 1e-04.
[1] "test RMSE of casual count prediction"
[1] 345.4918
[1] "model on registered model"
[1] "registered count"
Elasticnet

500 samples
 29 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 451, 450, 449, 450, 449, 450, ...
Resampling results across tuning parameters:

lambda	fraction	RMSE	Rsquared	MAE
0e+00	0.050	869.4870	0.6092200	714.5591
0e+00	0.525	494.4995	0.8324754	379.7170
0e+00	1.000	481.6494	0.8375684	362.1400

```
1e-04  0.050    1089.8830  0.4574527  901.2973
1e-04  0.525     533.4634  0.8151820  424.0403
1e-04  1.000     481.6357  0.8375767  362.1313
1e-01  0.050    1106.2667  0.4154972  913.8777
1e-01  0.525     590.2589  0.7912315  483.8361
1e-01  1.000     488.5703  0.8338632  370.7218
```

RMSE was used to select the optimal model using the smallest value.

The final values used for the model were fraction = 1 and lambda = 1e-04.

[1] "test RMSE of registered prediction"

[1] 896.3894

[1] "Test RMSE on Total count(casual +registered)"

[1] 1842.288

[1] "rpart"

[1] " model on casual count"

CART

500 samples

 30 predictor

No pre-processing

Resampling: Cross-Validated (10 fold, repeated 6 times)

Summary of sample sizes: 451, 450, 448, 451, 450, 451, ...

Resampling results across tuning parameters:

```
  cp          RMSE      Rsquared   MAE
  0.02492299  341.4464  0.6594160  236.1094
  0.02887968  347.6229  0.6441928  244.4587
  0.32414292  479.7115  0.5732009  353.8031
```

RMSE was used to select the optimal model using the smallest value.

The final value used for the model was cp = 0.02492299.

[1] "test RMSE of casual count prediction"

[1] 449.8056

[1] "model on registered model"

[1] "registered count"

CART

500 samples

 29 predictor

No pre-processing

Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 451, 450, 450, 450, 449, 448, ...
Resampling results across tuning parameters:

```
 cp        RMSE       Rsquared   MAE
 0.1169177  934.7520  0.3953798  761.8839
 0.1243528  974.0045  0.3453282  795.0006
 0.3125103  1118.7767  0.2190088  922.2833
```

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was cp = 0.1169177.
[1] "test RMSE of registered prediction"
[1] 1529.423
[1] "Test RMSE on Total count(casual +registered)"
[1] 2313.237

Python results:
1.model_results(Ridge_model)
test-RMSE of casual user count

305.360522612
coefficient of determination R^2 of the prediction

0.768492435934
test-RMSE of registered user count

539.04674324
coefficient of determination R^2 of the prediction

0.875461206306
RMSE of total count(registered+casual)
699.856823931
2.model_results(lasso_model)
test-RMSE of casual user count

303.540496172
coefficient of determination R^2 of the prediction

0.771243899729
test-RMSE of registered user count

539.262937572
coefficient of determination R^2 of the prediction

0.875361289254
RMSE of total count(registered+casual)
697.458102753

3.model_results(lin_reg_model)

test-RMSE of casual user count

306.027355775

coefficient of determination R^2 of the prediction

0.767480219418

test-RMSE of registered user count

540.929219525

coefficient of determination R^2 of the prediction

0.87458985075

RMSE of total count(registered+casual)

703.99924603

4. model_results(rf_model)

test-RMSE of casual user count

240.452837371

coefficient of determination R^2 of the prediction

0.856451315399

test-RMSE of registered user count

476.859579145

coefficient of determination R^2 of the prediction

0.90253855995

RMSE of total count(registered+casual)

572.848588392

5.model_results(DT_model)

test-RMSE of casual user count

361.070453267

coefficient of determination R^2 of the prediction

0.676314453591

test-RMSE of registered user count

576.441593842

coefficient of determination R^2 of the prediction

0.857582806781

RMSE of total count(registered+casual)

687.942621154

PCA results

R

[1] "Princiapl component analysis"

[1] "ridge"

[1] " model on casual count"

Ridge Regression

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 495, 495, 495, 496, 496, 495, ...
Resampling results across tuning parameters:

```
 lambda  RMSE         Rsquared   MAE
 0e+00   3.932391e-06  1.0000000  3.126141e-06
 1e-04   5.457959e-03  1.0000000  4.261238e-03
 1e-01   4.903872e+00  0.9999542  3.831106e+00
```

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was lambda = 0.

[1] "test RMSE of casual count prediction"

[1] 4.393815e-06

[1] "model on registered model"

[1] "registered count"

Ridge Regression

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 495, 494, 494, 496, 496, 495, ...
Resampling results across tuning parameters:

```
 lambda  RMSE         Rsquared   MAE
 0e+00   2.732269e-06  1.0000000  2.126584e-06
```

1e-04 1.194283e-02 1.0000000 9.163671e-03
1e-01 1.072713e+01 0.9999523 8.238587e+00

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was lambda = 0.
[1] "test RMSE of registered prediction"
[1] 3.5235e-06
[1] "Test RMSE on Total count(casual +registered)"
[1] 7.455615e-06

[1] "Princiapl component analysis"
[1] "lm"
[1] " model on casual count"
Linear Regression

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 495, 495, 494, 495, 495, 495, ...
Resampling results:

  RMSE   Rsquared MAE
  3.917672e-06 1   3.119728e-06

Tuning parameter 'intercept' was held constant at a value of TRUE
[1] "test RMSE of casual count prediction"
[1] 4.393815e-06
[1] "model on registered model"
[1] "registered count"
Linear Regression

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 494, 494, 496, 495, 495, 494, ...

Resampling results:

RMSE	Rsquared	MAE
2.740865e-06	1	2.137358e-06

Tuning parameter 'intercept' was held constant at a value of TRUE
[1] "test RMSE of registered prediction"
[1] 3.523501e-06
[1] "Test RMSE on Total count(casual +registered)"
[1] 7.455616e-06

[1] "Princiapl component analysis"
[1] "lasso"
[1] " model on casual count"
The lasso

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 496, 494, 495, 495, 495, 494, ...
Resampling results across tuning parameters:

fraction	RMSE	Rsquared	MAE
0.1	554.35104	1	428.45249
0.5	307.97231	1	238.02878
0.9	61.59357	1	47.60506

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was fraction = 0.9.
[1] "test RMSE of casual count prediction"
[1] 70.02412
[1] "model on registered model"
[1] "registered count"
The lasso

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 495, 494, 494, 494, 496, 495, ...
Resampling results across tuning parameters:

fraction	RMSE	Rsquared	MAE
0.1	1188.9124	1	960.9984
0.5	660.5067	1	533.8878
0.9	132.1010	1	106.7773

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was fraction = 0.9.
[1] "test RMSE of registered prediction"
[1] 229.2209
[1] "Test RMSE on Total count(casual +registered)"
[1] 267.4817

[1] "Princiapl component analysis"
[1] "glm"
[1] " model on casual count"
Generalized Linear Model

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 497, 494, 494, 495, 495, 495, ...
Resampling results:

RMSE	Rsquared	MAE
3.922186e-06	1	3.127059e-06

[1] "test RMSE of casual count prediction"
[1] 4.393815e-06
[1] "model on registered model"
[1] "registered count"

Generalized Linear Model

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 494, 494, 494, 496, 495, 494, …
Resampling results:

  RMSE   Rsquared MAE
  2.728588e-06 1   2.127038e-06

[1] "test RMSE of registered prediction"
[1] 3.523501e-06
[1] "Test RMSE on Total count(casual +registered)"
[1] 7.455616e-06

[1] "Princiapl component analysis"
[1] "enet"
[1] " model on casual count"
Elasticnet

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 495, 494, 495, 495, 494, 496, …
Resampling results across tuning parameters:

lambda	fraction	RMSE	Rsquared	MAE
0e+00	0.050	5.854962e+02	1.0000000	4.522145e+02
0e+00	0.525	2.927476e+02	1.0000000	2.261068e+02
0e+00	1.000	3.935200e-06	1.0000000	3.128242e-06
1e-04	0.050	5.854954e+02	1.0000000	4.522139e+02
1e-04	0.525	2.927393e+02	1.0000000	2.261004e+02
1e-04	1.000	5.459708e-03	1.0000000	4.218047e-03
1e-01	0.050	5.847381e+02	1.0000000	4.516291e+02
1e-01	0.525	2.847877e+02	1.0000000	2.199604e+02
1e-01	1.000	4.905737e+00	0.9999548	3.794031e+00

RMSE was used to select the optimal model using the smallest value.
The final values used for the model were fraction = 1 and lambda = 0.
[1] "test RMSE of casual count prediction"
[1] 4.393815e-06
[1] "model on registered model"
[1] "registered count"
Elasticnet

550 samples
 25 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 6 times)
Summary of sample sizes: 495, 495, 494, 495, 495, 495, ...
Resampling results across tuning parameters:

lambda	fraction	RMSE	Rsquared	MAE
0e+00	0.050	1.255152e+03	1.0000000	1.014410e+03
0e+00	0.525	6.275757e+02	1.0000000	5.072049e+02
0e+00	1.000	2.727754e-06	1.0000000	2.122407e-06
1e-04	0.050	1.255150e+03	1.0000000	1.014409e+03
1e-04	0.525	6.275572e+02	1.0000000	5.071900e+02
1e-04	1.000	1.178293e-02	1.0000000	8.941161e-03
1e-01	0.050	1.253543e+03	1.0000000	1.013111e+03
1e-01	0.525	6.106795e+02	1.0000000	4.935670e+02
1e-01	1.000	1.058525e+01	0.9999543	8.040639e+00

RMSE was used to select the optimal model using the smallest value. The final values used for the model were fraction = 1 and lambda = 0.

[1] "test RMSE of registered prediction"

[1] 3.5235e-06

[1] "Test RMSE on Total count(casual +registered)"

[1] 7.455615e-06

PCA results
Python

pca_model_results(Ridge_model)

test-RMSE PCA model

6.87384334901e-06

coefficient of determination R^2 of the prediction

1.0

test-RMSE PCA model

4.4653597263e-06

coefficient of determination R^2 of the prediction

1.0

RMSE of total count(registered+casual)

9.22966398454e-06

pca_model_results(lin_reg_model)

test-RMSE PCA model

4.2846475917e-06

coefficient of determination R^2 of the prediction

1.0

test-RMSE PCA model

3.17754368194e-06

coefficient of determination R^2 of the prediction

1.0

RMSE of total count(registered+casual)

6.11408842905e-06

pca_model_results(lasso_model)

test-RMSE PCA model

0.00159008940721

coefficient of determination R^2 of the prediction

0.999999999994

test-RMSE PCA model

0.000652521213658

coefficient of determination R^2 of the prediction

1.0

RMSE of total count(registered+casual)

0.00198170567784

pca_model_results(ela_net)

test-RMSE PCA model

0.00159008721435

coefficient of determination R^2 of the prediction

0.999999999994

test-RMSE PCA model

0.000652521066928

coefficient of determination R^2 of the prediction

1.0

RMSE of total count(registered+casual)

0.00198170349383

pca_model_results(rf_model)

test-RMSE PCA model

11.5971069189

coefficient of determination R^2 of the prediction

0.999666082848

test-RMSE PCA model

43.9180056722

coefficient of determination R^2 of the prediction

0.999173320777

RMSE of total count(registered+casual)

48.4319017912

Comparing results of the models
Results in r(feature selection using 'Boruta')

Model	R-square of casual count	R-square of registered count	Test RMSE of total predictions (casual + registered)
Linear regression	0.7276002	0.8383362	1855.027
Ridge regression	0.7310268	0.838133	1854.948
Lasso regression	0.7311360	0.8382036	1855.027
Elastic net regression	0.7314965	0.8400089	1854.972
GLM	0.7315754	0.8380084	1855.027

PCA results

Model	R-square of casual count	R-square of registered count	Test RMSE of total predictions (casual + registered)= count
Linear regression	1	1	7.455616e-06
Ridge regression	1	1	7.455615e-06
Lasso regression	1	1	267.4817
Elastic net regression	1	1	7.455615e-06
GLM	1	1	7.455616e-06

model results in python

Model	R-square of casual count	R-square of registered count	Test RMSE of total predictions (casual + registered)=

			count
Linear regression	0.767480219418	0.87458985075	703.99924603
Ridge regression	0.768492435934	0.875461206306	699.856823931
Lasso regression	0.771243899729	0.875361289254	697.458102753
Elastic net regression	0.58634102819	0.718459994794	1002.89775988
Decision tree	0.675011997498	0.864242707054	688.521999971
Random forest	0.827661295807	0.893328818247	591.324006456

PCA results in python

Model	R-square of casual count	R-square of registered count	Test RMSE of total predictions (casual + registered)= count
Linear regression	1	1	6.1140805e-06
Ridge regression	1	1	9.2296639e-06
Lasso regression	1	1	0.0019817056
Elastic net regression	1	1	0.0019817034
Decision tree	0.9989185533	0.999138188512	50.7063988949
Random forest	0.99961748340	0.999315745056	44.0029658091

So over all PCA has produced the best results for the above problem

NOTE:RMSE that I have provided in the above tables is of *** test data***.
RMSE of test data is less than the training data (the PCA models has achieved good results and it has not overfitted) with r_squared value of 100%.