

```

1  # -*- coding: utf-8 -*-
2  """
3  Created on Mon Mar 12 19:02:49 2018
4
5  @author: cml
6  """
7
8  import matplotlib.pyplot as plt
9  import numpy as np
10 from sklearn import datasets, linear_model
11 from sklearn.metrics import mean_squared_error, r2_score
12 import numpy as np
13 import pandas as pd
14 import statsmodels.api as sm
15 import statsmodels.formula.api as smf
16 from patsy import dmatrices
17 from sklearn.model_selection import LeaveOneOut
18 from sklearn import metrics
19
20 # Courtesy: https://codeburst.io/cross-validation-calculating
21 #           -r%C2%B2-and-accuracy-scores-after-loocv-5bd1015a50ec
22
23 df = pd.read_csv('Auto.csv', usecols=range(1,10))
24
25 X = df["horsepower"].values.reshape(-1,1) # our independent variable
26 y = df["mpg"].values.reshape(-1,1) # our dependent variable
27
28 loo = LeaveOneOut()
29 print('Splits: ', loo.get_n_splits(X))
30
31 #Arrays to store test data and predictions for each run
32 ytests = []
33 ypreds = []
34
35 #for each LOOCV split in X, fit a model using current x_train
36 #and y_train value. Save the array
37 for train_index, test_index in loo.split(X):
38     X_train, X_test = X[train_index], X[test_index]
39     y_train, y_test = y[train_index], y[test_index]
40
41     model = linear_model.LinearRegression()
42     model.fit(X = X_train, y = y_train)
43     y_pred = model.predict(X_test)
44
45     ytests += list(y_test)
46     ypreds += list(y_pred)
47
48 rr = metrics.r2_score(ytests, ypreds)
49 ms_error = metrics.mean_squared_error(ytests, ypreds)
50
51 print("LOOCV results:")
52 print("R^2: {:.5f}%, MSE: {:.5f}".format(rr*100, ms_error))

```