Projects in Statistical Learning

01 - Getting Started with Regression

Spencer Grewe

2023-07-16

Table of Contents

```
import pandas as pd
import numpy as np
import sklearn
from sklearn import linear_model
```

```
data = pd.read_csv('..\\Data_common\\student_mat.csv', sep = ';')
```

```
data.columns
```

```
Index(['school', 'sex', 'age', 'address', 'famsize', 'Pstatus', 'Medu', 'Fedu',
       'Mjob', 'Fjob', 'reason', 'guardian', 'traveltime', 'studytime',
       'failures', 'schoolsup', 'famsup', 'paid', 'activities', 'nursery',
      'higher', 'internet', 'romantic', 'famrel', 'freetime', 'goout', 'Dalc',
       'Walc', 'health', 'absences', 'G1', 'G2', 'G3'],
      dtype='object')
```

```
dat_copy = data[["G1", "G2", "G3", "studytime", "failures", "absences"]]
predict = "G3"

X = np.array(dat_copy.drop(columns = [predict])) # Features
y = np.array(dat_copy[predict]) # Labels
X.shape
```

```
(395, 5)
```

```
x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(X, y, test_

linearmodel = linear_model.LinearRegression()
linearmodel.fit(x_train, y_train)
acc = linearmodel.score(x_test, y_test)
print(acc)
print('Coefficient: \n', linearmodel.coef_) # These are each slope value
print('Intercept: \n', linearmodel.intercept_) # This is the intercept
```

```
0.9138665134600124
Coefficient:
 [ 0.14639984  0.97818194 -0.23267009 -0.32980368  0.03912193]
Intercept:
 -1.3227796844930353
```

```
predicts = linearmodel.predict(x_test)
for x in range(len(predicts)):
    print(predicts[x], x_test[x], y_test[x])
```

```
6.1229408627501085 [ 7  6  2  0 26] 6
12.040012736892848 [10 12  2  0 16] 11
9.173622877560025 [11  9  2  0 14] 9
12.01777547023882 [12 12  1  0  2] 11
9.76861190658997 [10 10  1  0  2] 10
8.343204060756216 [8 9 2 0 4] 10
11.941593104329488 [12 12  2  0  6] 12
8.333116175100487 [9 9 2 0 0] 10
19.27008153902225 [19 19  4  0  4] 20
9.012456301720412 [10  9  3  1 28] 9
14.266939033887668 [14 14  1  0  2] 14
18.223743618449735 [18 18  4  0  6] 18
13.058148495840758 [14 13  3  0  8] 14
12.763287326802006 [12 13  2  0  2] 12
13.379150331161544 [13 13  2  0 14] 14
4.431762322485655 [ 6  5  1  1 14] 5
15.762564350660181 [17 15  1  0  4] 16
3.884055293722298 [6 5 1 1 0] 0
5.290151226556218 [7 7 2 3 5] 7
```

```
7.6154407264765 74 [7 9 2 2 6] 8
15.080606866047038 [15 15  2  0  0] 15
10.719877731576481 [11 11  1  1  6] 10
12.029924851237123 [11 12  2  0 12] 11
13.476340351727952 [15 13  2  0  9] 15
11.913390942704881 [15 12  3  0  0] 14
19.899935828665523 [18 19  1  0  6] 19
16.283432509123372 [16 16  2  0  2] 17
12.598773213985037 [13 13  3  0  0] 13
12.161635198783342 [10 13  4  0  6] 13
14.266939033887668 [14 14  1  0  2] 13
12.872626728366715 [13 13  3  0  7] 14
12.779672557761113 [15 13  3  2 14] 13
6.150967379654692 [7 8 2 3 2] 9
14.805921468319738 [11 15  1  0  2] 15
14.934207029198948 [14 15  2  0  0] 15
7.725977765072239 [11  8  2  0  2] 8
4.487712487445706 [6 5 1 0 7] 6
16.35158848471955 [17 16  2  0  0] 16
9.576216111110073 [ 8  9  1  1 38] 8
13.809625247374521 [13 14  2  0  0] 15
```