math456 hw6

Luke Geel

March 2022

1 LASSO Regression for Diabetes Data Analysis

This code is written in python:

```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import datasets
from sklearn.model_selection import train_test_split, cross_val_score
diabetes = datasets.load_diabetes()
print(diabetes.DESCR)
```

Here I'm creating a dataframe from the dataset

```
df = pd.DataFrame(diabetes.data, columns = diabetes.feature_names)
df['measure'] = diabetes.target
df.head()
```

Finding relationship between BMI and target measurement by creating a scatterplot

```
x = df['bmi']
y = df['measure']
plt.scatter(x, y, marker='o')
plt.xlabel('bmi')
plt.ylabel('measure')
```

Preparing for linear regression df.head()

```
X = df.iloc[:,0:10]
X.head()
y = df.iloc[:,10]
y.head()
```

Split the set into training and test sets

```
from sklearn.linear_model import LinearRegression
```

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state=0)
lr = LinearRegression()
lr.fit(X_train, y_train)
y_pred = lr.predict(X_test)

Now I'm going to get coeffiecients for the linear model
dict_coef = dict(zip(diabetes.feature_names,lr.coef_))
dict_coef['Intercept'] = lr.intercept_
print(dict_coef)

Calculating R^2
from sklearn.metrics import r2_score
r2_score = r2_score(y_test, y_pred)
print("R2 score", r2_score)