```
 1 import numpy as np
 2 import pandas as pd
 3 import scipy.stats as stats
 4 from matplotlib import pyplot as plt
 5 from sklearn.linear_model import LinearRegression,
   Lasso, LassoCV
 6 from sklearn.model_selection import train_test_split
 7
 8 df=pd.read_csv('ML_HW_Data_Patients.csv')
 9 df.info
10
11
12 df['Height']=stats.zscore(df['Height'])
13 df['Weight']=stats.zscore(df['Weight'])
14 df=df.drop(['Diastolic','LastName'], axis=1)
15 df
16
17 df['Age']=stats.zscore(df['Age'])
18 Gender_dummy=pd.get_dummies(df.Gender)
19 Smoker_dummy=pd.get_dummies(df.Smoker)
20 Location_dummy=pd.get_dummies(df.Location)
21 Self_dummy=pd.get_dummies(df.SelfAssessedHealthStatus
   )
22
23 new_df=pd.concat([df,Gender_dummy,Smoker_dummy,
   Location_dummy,Self_dummy],axis=1)
24 new_df.drop(labels=['Gender','Smoker','Location',"'
   Female'",0,'SelfAssessedHealthStatus',
25                     "'St. Mary's Medical Center'","'
   Poor'"],axis=1,inplace=True)
26 new_df.rename(columns={1:'Smoker',"'Male'":'Male',"'
   county General Hospital'":'countyGenralHospital', "'
   VA Hospital'":'VAHospital',"'Excellent'":'Excellent',
27                         "'Fair'":'Fair',"'Good'":'Good
   '},inplace=True)
28
29
30 X= new_df.drop(['Systolic'],axis=1)
31 Y=new_df['Systolic']
32
33 X_train,X_test,y_train,y_test=train_test_split(X,Y,
```

```
33 test_size=0.033,random_state=42)
34 reg=LinearRegression()
35 reg.fit(X_train,y_train)
36 y_predict=reg.predict(X_test)
37
38 model=LassoCV(eps=1e-3,cv=10)
39 model.fit(X,Y)
40 alphas=model.alphas_
41 alpha=model.alpha_
42 mse_path=model.mse_path_
43
44
45 print('alpha:',model.alpha_)
46 print('Intercept:',model.intercept_)
47 print('Coefficients:',model.coef_)
48
49
50 coefs=[]
51 for a in alphas:
52     clf=Lasso(alpha=a)
53     clf.fit(X,Y)
54     coefs.append(clf.coef_)
55 coefs_path=np.mat(coefs)
56
57
58 plt.plot(-np.log10(model.alphas_),coefs_path,
   linestyle='solid')
59 ymin,ymax=plt.ylim()
60 plt.vlines(-np.log10(model.alpha_),ymin,ymax,
   linestyle='dotted',label='Selected alpha')
61 plt.legend()
62 plt.xlabel('-log(alpha)')
63 plt.ylabel('Regression Coefficients')
64 plt.title('Lasso plot')
65 plt.show()
66
67 plt.plot(-np.log10(model.alphas_),mse_path, ':')
68 plt.plot(-np.log10(model.alphas_),mse_path.mean(axis
   =-1),10)
69 ymin,ymax=0,60
70 plt.vlines(-np.log10(model.alpha_),ymin,ymax,
```

```
70 linestyle='dotted',label='Selected alpha')
71 plt.legend()
72 plt.xlabel('-log(alpha)')
73 plt.ylabel('Mean square error')
74 plt.title('mean Square Errors on each CV fold')
75 plt.ylim(ymin,ymax)
76 plt.show()
77
78 la={}
79 la_df=pd.DataFrame(la)
80 i=0
81 for a in alphas:
82     choose=Lasso(alpha=a)
83     choose.fit(X,Y)
84     la[i]=[a,choose.coef_,choose.intercept_]
85     i+=1
86 print('\n',la)
87
88
89 print('Coffients:',Lasso(alpha=0.9021926169511577).
   fit(X,Y).coef_)
```