breastcancer

May 20, 2024

[ ]:
```
#  loadding the necessary libraries
from dataidea.packages import *
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.linear_model import Lasso
from sklearn import metrics
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
```

[ ]:
```
# loading the dataset
breast_cancer_dataset = pd.read_csv('datasets/breast-cancer.csv')
```

[ ]:
```
# checking the dataset
breast_cancer_dataset
```

[ ]:
```
           id diagnosis  radius_mean  texture_mean  perimeter_mean  area_mean  \
0      842302         M        17.99         10.38          122.80     1001.0
1      842517         M        20.57         17.77          132.90     1326.0
2    84300903         M        19.69         21.25          130.00     1203.0
3    84348301         M        11.42         20.38           77.58      386.1
4    84358402         M        20.29         14.34          135.10     1297.0
..        …       …          …             …               …          …
564    926424         M        21.56         22.39          142.00     1479.0
565    926682         M        20.13         28.25          131.20     1261.0
566    926954         M        16.60         28.08          108.30      858.1
567    927241         M        20.60         29.33          140.10     1265.0
568     92751         B         7.76         24.54           47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave points_mean  \
0            0.11840           0.27760         0.30010              0.14710
1            0.08474           0.07864         0.08690              0.07017
2            0.10960           0.15990         0.19740              0.12790
3            0.14250           0.28390         0.24140              0.10520
4            0.10030           0.13280         0.19800              0.10430
..               …                 …               …                    …
564          0.11100           0.11590         0.24390              0.13890
```

```
565          0.09780           0.10340         0.14400              0.09791
566          0.08455           0.10230         0.09251              0.05302
567          0.11780           0.27700         0.35140              0.15200
568          0.05263           0.04362         0.00000              0.00000

     …  radius_worst  texture_worst  perimeter_worst  area_worst  \
0    …        25.380          17.33           184.60      2019.0
1    …        24.990          23.41           158.80      1956.0
2    …        23.570          25.53           152.50      1709.0
3    …        14.910          26.50            98.87       567.7
4    …        22.540          16.67           152.20      1575.0
..   …           …              …                …          …
564  …        25.450          26.40           166.10      2027.0
565  …        23.690          38.25           155.00      1731.0
566  …        18.980          34.12           126.70      1124.0
567  …        25.740          39.42           184.60      1821.0
568  …         9.456          30.37            59.16       268.6

     smoothness_worst  compactness_worst  concavity_worst  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                …                  …                …
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst
0                  0.2654          0.4601                  0.11890
1                  0.1860          0.2750                  0.08902
2                  0.2430          0.3613                  0.08758
3                  0.2575          0.6638                  0.17300
4                  0.1625          0.2364                  0.07678
..                    …               …                        …
564                0.2216          0.2060                  0.07115
565                0.1628          0.2572                  0.06637
566                0.1418          0.2218                  0.07820
567                0.2650          0.4087                  0.12400
568                0.0000          0.2871                  0.07039

[569 rows x 32 columns]
```

[ ]:
```
#  lets us predict the worst_fractional_dimension of the person
breast_cancer_dataset.head()
```

[ ]:

	id	diagnosis	radius_mean	texture_mean	perimeter_mean	area_mean	\
0	842302	M	17.99	10.38	122.80	1001.0	
1	842517	M	20.57	17.77	132.90	1326.0	
2	84300903	M	19.69	21.25	130.00	1203.0	
3	84348301	M	11.42	20.38	77.58	386.1	
4	84358402	M	20.29	14.34	135.10	1297.0	

	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	\
0	0.11840	0.27760	0.3001	0.14710	
1	0.08474	0.07864	0.0869	0.07017	
2	0.10960	0.15990	0.1974	0.12790	
3	0.14250	0.28390	0.2414	0.10520	
4	0.10030	0.13280	0.1980	0.10430	

	…	radius_worst	texture_worst	perimeter_worst	area_worst	\
0	…	25.38	17.33	184.60	2019.0	
1	…	24.99	23.41	158.80	1956.0	
2	…	23.57	25.53	152.50	1709.0	
3	…	14.91	26.50	98.87	567.7	
4	…	22.54	16.67	152.20	1575.0	

	smoothness_worst	compactness_worst	concavity_worst	concave points_worst	\
0	0.1622	0.6656	0.7119	0.2654	
1	0.1238	0.1866	0.2416	0.1860	
2	0.1444	0.4245	0.4504	0.2430	
3	0.2098	0.8663	0.6869	0.2575	
4	0.1374	0.2050	0.4000	0.1625	

	symmetry_worst	fractal_dimension_worst
0	0.4601	0.11890
1	0.2750	0.08902
2	0.3613	0.08758
3	0.6638	0.17300
4	0.2364	0.07678

[5 rows x 32 columns]

[ ]:
```
breast_cancer_dataset.shape
```

[ ]: (569, 32)

[ ]:
```
# checking out the diagonsis column
breast_cancer_dataset['diagnosis'].value_counts()
```

[ ]:
```
diagnosis
B    357
M    212
Name: count, dtype: int64
```

[ ]:
```
# relationship of diagonsis
breast_cancer_dataset.groupby('diagnosis').mean()
```

[ ]:
```
                     id  radius_mean  texture_mean  perimeter_mean  \
diagnosis
B          2.654382e+07    12.146524     17.914762       78.075406
M          3.681805e+07    17.462830     21.604906      115.365377

             area_mean  smoothness_mean  compactness_mean  concavity_mean  \
diagnosis
B           462.790196         0.092478          0.080085        0.046058
M           978.376415         0.102898          0.145188        0.160775

            concave points_mean  symmetry_mean  …  radius_worst  \
diagnosis                                       …
B                      0.025717       0.174186  …     13.379801
M                      0.087990       0.192909  …     21.134811

            texture_worst  perimeter_worst   area_worst  smoothness_worst  \
diagnosis
B               23.515070        87.005938   558.899440          0.124959
M               29.318208       141.370330  1422.286321          0.144845

            compactness_worst  concavity_worst  concave points_worst  \
diagnosis
B                    0.182673         0.166238              0.074444
M                    0.374824         0.450606              0.182237

            symmetry_worst  fractal_dimension_worst
diagnosis
B                 0.270246                 0.079442
M                 0.323468                 0.091530

[2 rows x 31 columns]
```

[ ]:
```
# assigning variables
x= breast_cancer_dataset.
  ↪drop(columns=['fractal_dimension_worst','diagnosis'],axis=1)
y=breast_cancer_dataset['fractal_dimension_worst']
```

[ ]:
```
# printing them out
x
```

[ ]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	\
0	842302	17.99	10.38	122.80	1001.0	
1	842517	20.57	17.77	132.90	1326.0	
2	84300903	19.69	21.25	130.00	1203.0	
3	84348301	11.42	20.38	77.58	386.1	
4	84358402	20.29	14.34	135.10	1297.0	
..	…	…	…	…	…	
564	926424	21.56	22.39	142.00	1479.0	
565	926682	20.13	28.25	131.20	1261.0	
566	926954	16.60	28.08	108.30	858.1	
567	927241	20.60	29.33	140.10	1265.0	
568	92751	7.76	24.54	47.92	181.0	

	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	\
0	0.11840	0.27760	0.30010	0.14710	
1	0.08474	0.07864	0.08690	0.07017	
2	0.10960	0.15990	0.19740	0.12790	
3	0.14250	0.28390	0.24140	0.10520	
4	0.10030	0.13280	0.19800	0.10430	
..	…	…	…	…	
564	0.11100	0.11590	0.24390	0.13890	
565	0.09780	0.10340	0.14400	0.09791	
566	0.08455	0.10230	0.09251	0.05302	
567	0.11780	0.27700	0.35140	0.15200	
568	0.05263	0.04362	0.00000	0.00000	

	symmetry_mean	…	fractal_dimension_se	radius_worst	texture_worst	\
0	0.2419	…	0.006193	25.380	17.33	
1	0.1812	…	0.003532	24.990	23.41	
2	0.2069	…	0.004571	23.570	25.53	
3	0.2597	…	0.009208	14.910	26.50	
4	0.1809	…	0.005115	22.540	16.67	
..	…	…	…	…	…	
564	0.1726	…	0.004239	25.450	26.40	
565	0.1752	…	0.002498	23.690	38.25	
566	0.1590	…	0.003892	18.980	34.12	
567	0.2397	…	0.006185	25.740	39.42	
568	0.1587	…	0.002783	9.456	30.37	

	perimeter_worst	area_worst	smoothness_worst	compactness_worst	\
0	184.60	2019.0	0.16220	0.66560	
1	158.80	1956.0	0.12380	0.18660	
2	152.50	1709.0	0.14440	0.42450	
3	98.87	567.7	0.20980	0.86630	
4	152.20	1575.0	0.13740	0.20500	
..	…	…	…	…	
564	166.10	2027.0	0.14100	0.21130	

```
565          155.00      1731.0           0.11660           0.19220
566          126.70      1124.0           0.11390           0.30940
567          184.60      1821.0           0.16500           0.86810
568           59.16       268.6           0.08996           0.06444

     concavity_worst  concave points_worst  symmetry_worst
0             0.7119                0.2654          0.4601
1             0.2416                0.1860          0.2750
2             0.4504                0.2430          0.3613
3             0.6869                0.2575          0.6638
4             0.4000                0.1625          0.2364
..               …                     …               …
564           0.4107                0.2216          0.2060
565           0.3215                0.1628          0.2572
566           0.3403                0.1418          0.2218
567           0.9387                0.2650          0.4087
568           0.0000                0.0000          0.2871

[569 rows x 30 columns]
```

[ ]:
```
y
```

[ ]:
```
0      0.11890
1      0.08902
2      0.08758
3      0.17300
4      0.07678
       …
564    0.07115
565    0.06637
566    0.07820
567    0.12400
568    0.07039
Name: fractal_dimension_worst, Length: 569, dtype: float64
```

[ ]:
```
# train_test_split
x_train,x_test,y_train,y_test= train_test_split(x , y, test_size=0.2,␣
 ↪random_state=42)
```

[ ]:
```
# checking out the shapes
x_train.shape,x_test.shape,y_train.shape
```

[ ]:
```
((455, 30), (114, 30), (455,))
```

[ ]:
```
x_train
```

[ ]:

	id	radius_mean	texture_mean	perimeter_mean	area_mean	\
68	859471	9.029	17.33	58.79	250.5	
181	873593	21.090	26.57	142.70	1311.0	
63	859196	9.173	13.86	59.20	260.9	
248	88466802	10.650	25.22	68.01	347.0	
60	858970	10.170	14.88	64.55	311.9	
..	…	…	…	…	…	
71	859711	8.888	14.64	58.79	244.0	
106	863031	11.640	18.33	75.17	412.5	
270	8910721	14.290	16.82	90.30	632.6	
435	908489	13.980	19.62	91.12	599.5	
102	862965	12.180	20.52	77.22	458.7	

	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	\
68	0.10660	0.14130	0.31300	0.04375	
181	0.11410	0.28320	0.24870	0.14960	
63	0.07721	0.08751	0.05988	0.02180	
248	0.09657	0.07234	0.02379	0.01615	
60	0.11340	0.08061	0.01084	0.01290	
..	…	…	…	…	
71	0.09783	0.15310	0.08606	0.02872	
106	0.11420	0.10170	0.07070	0.03485	
270	0.06429	0.02675	0.00725	0.00625	
435	0.10600	0.11330	0.11260	0.06463	
102	0.08013	0.04038	0.02383	0.01770	

	symmetry_mean	…	fractal_dimension_se	radius_worst	texture_worst	\
68	0.2111	…	0.009559	10.310	22.65	
181	0.2395	…	0.005295	26.680	33.48	
63	0.2341	…	0.005822	10.010	19.23	
248	0.1897	…	0.002619	12.250	35.19	
60	0.2743	…	0.005953	11.020	17.45	
..	…	…	…	…	…	
71	0.1902	…	0.021930	9.733	15.67	
106	0.1801	…	0.003840	13.140	29.26	
270	0.1508	…	0.001381	14.910	20.65	
435	0.1669	…	0.002846	17.040	30.80	
102	0.1739	…	0.001532	13.340	32.84	

	perimeter_worst	area_worst	smoothness_worst	compactness_worst	\
68	65.50	324.7	0.14820	0.43650	
181	176.50	2089.0	0.14910	0.75840	
63	65.59	310.1	0.09836	0.16780	
248	77.98	455.7	0.14990	0.13980	
60	69.86	368.6	0.12750	0.09866	
..	…	…	…	…	
71	62.56	284.4	0.12070	0.24360	

```
106            85.51       521.7           0.16880           0.26600
270            94.44       684.6           0.08567           0.05036
435           113.90       869.3           0.16130           0.35680
102            84.58       547.8           0.11230           0.08862

     concavity_worst  concave points_worst  symmetry_worst
68           1.25200               0.17500          0.4228
181          0.67800               0.29030          0.4098
63           0.13970               0.05087          0.3282
248          0.11250               0.06136          0.3409
60           0.02168               0.02579          0.3557
..               …                     …               …
71           0.14340               0.04786          0.2254
106          0.28730               0.12180          0.2806
270          0.03866               0.03333          0.2458
435          0.40690               0.18270          0.3179
102          0.11450               0.07431          0.2694

[455 rows x 30 columns]
```

[ ]:
```
#  our model fitting
model = LinearRegression()

# fitting the model
model.fit(x_train,y_train)
```

[ ]:
```
LinearRegression()
```

[ ]:
```
#  Evaluating our model
y_predicted_values = model.predict(x_train)
x_train_accuracy = metrics.r2_score(y_predicted_values,y_train)
```

[ ]:
```
# checking out the accuracy
print('The accuracy score is:', x_train_accuracy)
```

```
The accuracy score is: 0.9446741041971212
```

[ ]:
```
y_predicted_values = model.predict(x_train)
mean_accuracy = metrics.mean_absolute_error(y_predicted_values,y_train)
```

[ ]:
```
print('The accuracy score is:', mean_accuracy)
```

```
The accuracy score is: 0.0028635781048222306
```

[ ]:
```
#  Evaluating the x_test
y_predicted_= model.predict(x_test)
x_test_accuracy = metrics.r2_score(y_predicted_,y_test)
```

[ ]:
```
# printing out the accuracy results on x_test
print('the accuracy score is : ' , x_test_accuracy)
```

```
the accuracy score is :  0.9265111011945508
```

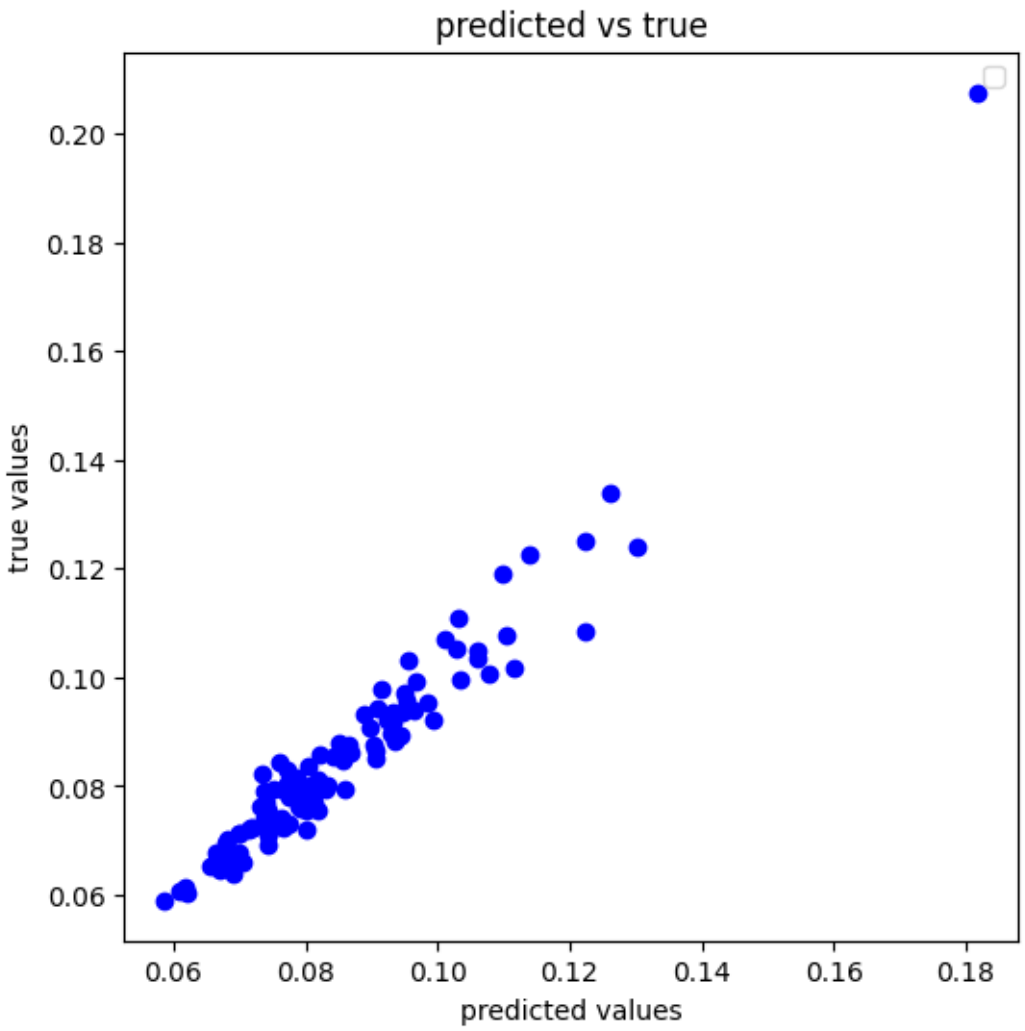
[ ]:
```
#  Evaluating the x_test
y_predicted_= model.predict(x_test)
mean_absolte= metrics.mean_absolute_error(y_predicted_,y_test)
```

[ ]:
```
print('the accuracy score is : ' , mean_absolte)
```

```
the accuracy score is :  0.0032648439120596274
```

[ ]:
```
# plotting to see how close the true is from the predicted
plt.figure(figsize=(6,6))
plt.scatter(y_predicted_,y_test,color='blue')
plt.title('predicted vs true')
plt.xlabel('predicted values')
plt.ylabel('true values')
plt.legend()
plt.show()
```

```
No artists with labels found to put in legend.  Note that artists whose label
start with an underscore are ignored when legend() is called with no argument.
```



[ ]:
```
#  our model fitting
lasso_reg = Lasso()

# fitting the model
lasso_reg.fit(x_train,y_train)
```

[ ]:
```
Lasso()
```

[ ]:
```
#  Evaluating our model
y_predicted_values = lasso_reg.predict(x_train)
mean_absolute_error = metrics.mean_absolute_error(y_predicted_values,y_train)
```

[ ]:
```
# checking out the accuracy
print('The accuracy score is:', mean_absolute_error)
```

```
The accuracy score is: 0.013501274711328828
```

[ ]:
```
#  Encoding practice
Label_encoder = LabelEncoder()
labels = Label_encoder.fit_transform(breast_cancer_dataset['diagnosis'])
breast_cancer_dataset['target']=labels
breast_cancer_dataset.drop(columns='diagnosis' , axis=1 ,inplace=True)
```

[ ]:
```
breast_cancer_dataset
```

[ ]:
```
           id  radius_mean  texture_mean  perimeter_mean  area_mean  \
0      842302        17.99         10.38          122.80     1001.0
1      842517        20.57         17.77          132.90     1326.0
2    84300903        19.69         21.25          130.00     1203.0
3    84348301        11.42         20.38           77.58      386.1
4    84358402        20.29         14.34          135.10     1297.0
..        …           …             …               …          …
564    926424        21.56         22.39          142.00     1479.0
565    926682        20.13         28.25          131.20     1261.0
566    926954        16.60         28.08          108.30      858.1
567    927241        20.60         29.33          140.10     1265.0
568     92751         7.76         24.54           47.92      181.0

     smoothness_mean  compactness_mean  concavity_mean  concave points_mean  \
0            0.11840           0.27760         0.30010              0.14710
1            0.08474           0.07864         0.08690              0.07017
2            0.10960           0.15990         0.19740              0.12790
3            0.14250           0.28390         0.24140              0.10520
4            0.10030           0.13280         0.19800              0.10430
..               …                 …               …                    …
564          0.11100           0.11590         0.24390              0.13890
565          0.09780           0.10340         0.14400              0.09791
566          0.08455           0.10230         0.09251              0.05302
567          0.11780           0.27700         0.35140              0.15200
568          0.05263           0.04362         0.00000              0.00000

     symmetry_mean  …  texture_worst  perimeter_worst  area_worst  \
0           0.2419  …          17.33           184.60      2019.0
1           0.1812  …          23.41           158.80      1956.0
2           0.2069  …          25.53           152.50      1709.0
3           0.2597  …          26.50            98.87       567.7
4           0.1809  …          16.67           152.20      1575.0
..             …  …              …                …          …
564         0.1726  …          26.40           166.10      2027.0
565         0.1752  …          38.25           155.00      1731.0
566         0.1590  …          34.12           126.70      1124.0
567         0.2397  …          39.42           184.60      1821.0
```

```
568         0.1587  …          30.37            59.16       268.6

     smoothness_worst  compactness_worst  concavity_worst  \
0             0.16220            0.66560           0.7119
1             0.12380            0.18660           0.2416
2             0.14440            0.42450           0.4504
3             0.20980            0.86630           0.6869
4             0.13740            0.20500           0.4000
..                …                  …                …
564           0.14100            0.21130           0.4107
565           0.11660            0.19220           0.3215
566           0.11390            0.30940           0.3403
567           0.16500            0.86810           0.9387
568           0.08996            0.06444           0.0000

     concave points_worst  symmetry_worst  fractal_dimension_worst  target
0                  0.2654          0.4601                  0.11890       1
1                  0.1860          0.2750                  0.08902       1
2                  0.2430          0.3613                  0.08758       1
3                  0.2575          0.6638                  0.17300       1
4                  0.1625          0.2364                  0.07678       1
..                    …               …                        …       …
564                0.2216          0.2060                  0.07115       1
565                0.1628          0.2572                  0.06637       1
566                0.1418          0.2218                  0.07820       1
567                0.2650          0.4087                  0.12400       1
568                0.0000          0.2871                  0.07039       0

[569 rows x 32 columns]
```