breast_cancer_xgboost

December 5, 2020

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[2]:
```
data = pd.read_csv("breast_cancer_data.csv")
```

[3]:
```
data.head()
```

[3]:

	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	

	…	texture_worst	perimeter_worst	area_worst	smoothness_worst	\
0	…	17.33	184.60	2019.0	0.1622	
1	…	23.41	158.80	1956.0	0.1238	
2	…	25.53	152.50	1709.0	0.1444	
3	…	26.50	98.87	567.7	0.2098	
4	…	16.67	152.20	1575.0	0.1374	

	compactness_worst	concavity_worst	concave points_worst	symmetry_worst	\
0	0.6656	0.7119	0.2654	0.4601	
1	0.1866	0.2416	0.1860	0.2750	
2	0.4245	0.4504	0.2430	0.3613	
3	0.8663	0.6869	0.2575	0.6638	
4	0.2050	0.4000	0.1625	0.2364	

```
   fractal_dimension_worst  Unnamed: 32
0                  0.11890          NaN
1                  0.08902          NaN
2                  0.08758          NaN
3                  0.17300          NaN
4                  0.07678          NaN

[5 rows x 33 columns]
```

[4]:
```
data.columns
```

[4]:
```
Index(['id', 'diagnosis', 'radius_mean', 'texture_mean', 'perimeter_mean',
       'area_mean', 'smoothness_mean', 'compactness_mean', 'concavity_mean',
       'concave points_mean', 'symmetry_mean', 'fractal_dimension_mean',
       'radius_se', 'texture_se', 'perimeter_se', 'area_se', 'smoothness_se',
       'compactness_se', 'concavity_se', 'concave points_se', 'symmetry_se',
       'fractal_dimension_se', 'radius_worst', 'texture_worst',
       'perimeter_worst', 'area_worst', 'smoothness_worst',
       'compactness_worst', 'concavity_worst', 'concave points_worst',
       'symmetry_worst', 'fractal_dimension_worst', 'Unnamed: 32'],
      dtype='object')
```

[5]:
```
data = data.drop(['id'], axis = 1)
data = data.drop(['Unnamed: 32'], axis = 1)
```

[6]:
```
# Missing Value check
data.isnull().sum()
```

[6]:
```
diagnosis                 0
radius_mean               0
texture_mean              0
perimeter_mean            0
area_mean                 0
smoothness_mean           0
compactness_mean          0
concavity_mean            0
concave points_mean       0
symmetry_mean             0
fractal_dimension_mean    0
radius_se                 0
texture_se                0
perimeter_se              0
area_se                   0
smoothness_se             0
compactness_se            0
concavity_se              0
concave points_se         0
```

```
symmetry_se                0
fractal_dimension_se       0
radius_worst               0
texture_worst              0
perimeter_worst            0
area_worst                 0
smoothness_worst           0
compactness_worst          0
concavity_worst            0
concave points_worst       0
symmetry_worst             0
fractal_dimension_worst    0
dtype: int64
```

[7]:
```
# Summary values of the dataset
data.describe()
```

[7]:

	radius_mean	texture_mean	perimeter_mean	area_mean	\
count	569.000000	569.000000	569.000000	569.000000	
mean	14.127292	19.289649	91.969033	654.889104	
std	3.524049	4.301036	24.298981	351.914129	
min	6.981000	9.710000	43.790000	143.500000	
25%	11.700000	16.170000	75.170000	420.300000	
50%	13.370000	18.840000	86.240000	551.100000	
75%	15.780000	21.800000	104.100000	782.700000	
max	28.110000	39.280000	188.500000	2501.000000	

	smoothness_mean	compactness_mean	concavity_mean	concave points_mean	\
count	569.000000	569.000000	569.000000	569.000000	
mean	0.096360	0.104341	0.088799	0.048919	
std	0.014064	0.052813	0.079720	0.038803	
min	0.052630	0.019380	0.000000	0.000000	
25%	0.086370	0.064920	0.029560	0.020310	
50%	0.095870	0.092630	0.061540	0.033500	
75%	0.105300	0.130400	0.130700	0.074000	
max	0.163400	0.345400	0.426800	0.201200	

	symmetry_mean	fractal_dimension_mean	…	radius_worst	\
count	569.000000	569.000000	…	569.000000	
mean	0.181162	0.062798	…	16.269190	
std	0.027414	0.007060	…	4.833242	
min	0.106000	0.049960	…	7.930000	
25%	0.161900	0.057700	…	13.010000	
50%	0.179200	0.061540	…	14.970000	
75%	0.195700	0.066120	…	18.790000	
max	0.304000	0.097440	…	36.040000	

```
       texture_worst  perimeter_worst   area_worst  smoothness_worst  \
count     569.000000       569.000000   569.000000        569.000000
mean       25.677223       107.261213   880.583128          0.132369
std         6.146258        33.602542   569.356993          0.022832
min        12.020000        50.410000   185.200000          0.071170
25%        21.080000        84.110000   515.300000          0.116600
50%        25.410000        97.660000   686.500000          0.131300
75%        29.720000       125.400000  1084.000000          0.146000
max        49.540000       251.200000  4254.000000          0.222600

       compactness_worst  concavity_worst  concave points_worst  \
count         569.000000       569.000000            569.000000
mean            0.254265         0.272188              0.114606
std             0.157336         0.208624              0.065732
min             0.027290         0.000000              0.000000
25%             0.147200         0.114500              0.064930
50%             0.211900         0.226700              0.099930
75%             0.339100         0.382900              0.161400
max             1.058000         1.252000              0.291000

       symmetry_worst  fractal_dimension_worst
count      569.000000               569.000000
mean         0.290076                 0.083946
std          0.061867                 0.018061
min          0.156500                 0.055040
25%          0.250400                 0.071460
50%          0.282200                 0.080040
75%          0.317900                 0.092080
max          0.663800                 0.207500

[8 rows x 30 columns]
```

[30]:
```
import boto3, re, sys, math, json, os, sagemaker, urllib.request
from sagemaker import get_execution_role
from sagemaker.session import TrainingInput
from sagemaker.serializers import CSVSerializer
```

[9]:
```
# Defining IAM Role and setting up S3 bucket
role = get_execution_role()
prefix = 'sagemaker/breastcancer' # Folder to be created inside bucket
bucket_name = sagemaker.Session().default_bucket()
my_region = boto3.session.Session().region_name
print(my_region)
```

```
us-east-1
```

[10]:
```
bucket_name
```

[10]:
```
'sagemaker-us-east-1-844927434406'
```

[11]:
```
# Train-test split
train_data, test_data = np.split(data.sample(frac=1, random_state=123), [int(0.
↪8 * len(data))])
```

[61]:
```
train_data.diagnosis = pd.Categorical(pd.factorize(train_data.diagnosis)[0])
test_data.diagnosis = pd.Categorical(pd.factorize(test_data.diagnosis)[0])
```

[62]:
```
# AWS Sagemaker requires data to be in csv format without headers
train_data.to_csv('train.csv', index = False, header = False)
```

[63]:
```
# Uploading training data to S3 bucket
boto3.Session().resource('s3').Bucket(bucket_name).Object(os.path.join(prefix,␣
↪'train/train.csv')).upload_file('train.csv')

# Setting up training data input path for sagemaker
s3_input_train = TrainingInput('s3://{}/{}/train/'.format(bucket_name, prefix),␣
↪content_type='text/csv')
```

[18]:
```
# Extreme Gradient Boosting
from sagemaker import image_uris

xgb_container = image_uris.retrieve("xgboost", my_region, "1.0-1")
sess = sagemaker.Session()
# Initialize the estimator
xgb = sagemaker.estimator.Estimator(image_uri = xgb_container,
                role=role,
                instance_count=1,
                instance_type='ml.m4.xlarge',
                output_path='s3://{}/{}/output'.format(bucket_name, prefix),
                sagemaker_session=sess)

# Set hyperparameters
xgb.set_hyperparameters(
        max_depth = 5,
        eta = 0.2,
        objective = "binary:logistic",
        num_round = 1000)
```

[64]:
```
# Training the model
# Ignore warning messages on output
xgb.fit({'train': s3_input_train})
```

```
2020-12-04 19:33:49 Starting - Starting the training job…
2020-12-04 19:33:51 Starting - Launching requested ML instances…
2020-12-04 19:35:06 Starting - Preparing the instances for training…
```

```
2020-12-04 19:36:00 Downloading - Downloading input data…
2020-12-04 19:36:26 Training - Downloading the training
image..INFO:sagemaker-containers:Imported framework

sagemaker_xgboost_container.training
INFO:sagemaker-containers:Failed to parse hyperparameter objective value

binary:logistic to Json.
Returning the value itself
INFO:sagemaker-containers:No GPUs detected (normal if no gpus

installed)
INFO:sagemaker_xgboost_container.training:Running XGBoost Sagemaker in

algorithm mode
INFO:root:Determined delimiter of CSV input is ','
INFO:root:Determined delimiter of CSV input is ','
[19:37:01] 455x30 matrix with 13650 entries loaded from

/opt/ml/input/data/train?format=csv&label_column=0&delimiter=,
INFO:root:Single node training.
INFO:root:Train matrix has 455 rows
[19:37:01] WARNING: /workspace/src/learner.cc:328:
Parameters: { num_round } might not be used.

  This may not be accurate due to some parameters are only used in language

bindings but

  passed down to XGBoost core.  Or some parameters are not used but slip through

this

  verification. Please open an issue if you find above cases.


[0]#011train-error:0.06593
[1]#011train-error:0.05934
[2]#011train-error:0.05495
[3]#011train-error:0.04176
[4]#011train-error:0.04615
[5]#011train-error:0.03736
[6]#011train-error:0.03077
[7]#011train-error:0.03077
[8]#011train-error:0.02857
[9]#011train-error:0.02637
[10]#011train-error:0.02418
[11]#011train-error:0.02418
[12]#011train-error:0.02418
[13]#011train-error:0.02198
[14]#011train-error:0.01978
[15]#011train-error:0.01758
[16]#011train-error:0.01978
```

```
[17]#011train-error:0.01319
[18]#011train-error:0.01538
[19]#011train-error:0.01538
[20]#011train-error:0.01978
[21]#011train-error:0.01758
[22]#011train-error:0.01978
[23]#011train-error:0.01758
[24]#011train-error:0.01538
[25]#011train-error:0.01319
[26]#011train-error:0.01319
[27]#011train-error:0.01099
[28]#011train-error:0.01099
[29]#011train-error:0.00879
[30]#011train-error:0.01319
[31]#011train-error:0.01099
[32]#011train-error:0.01099
[33]#011train-error:0.01099
[34]#011train-error:0.01319
[35]#011train-error:0.01319
[36]#011train-error:0.01319
[37]#011train-error:0.01099
[38]#011train-error:0.01099
[39]#011train-error:0.01099
[40]#011train-error:0.01099
[41]#011train-error:0.01099
[42]#011train-error:0.01099
[43]#011train-error:0.01099
[44]#011train-error:0.01099
[45]#011train-error:0.01099
[46]#011train-error:0.01099
[47]#011train-error:0.01099
[48]#011train-error:0.01319
[49]#011train-error:0.01319
[50]#011train-error:0.01319
[51]#011train-error:0.01319
[52]#011train-error:0.01319
[53]#011train-error:0.01319
[54]#011train-error:0.01319
[55]#011train-error:0.01319
[56]#011train-error:0.01319
[57]#011train-error:0.01319
[58]#011train-error:0.01319
[59]#011train-error:0.01319
[60]#011train-error:0.01319
[61]#011train-error:0.01319
[62]#011train-error:0.01319
[63]#011train-error:0.01319
[64]#011train-error:0.01319
```

[65]#011train-error:0.01319
[66]#011train-error:0.01319
[67]#011train-error:0.01319
[68]#011train-error:0.01319
[69]#011train-error:0.01319
[70]#011train-error:0.01319
[71]#011train-error:0.01319
[72]#011train-error:0.01319
[73]#011train-error:0.01319
[74]#011train-error:0.01319
[75]#011train-error:0.01319
[76]#011train-error:0.01319
[77]#011train-error:0.01319
[78]#011train-error:0.01319
[79]#011train-error:0.01319
[80]#011train-error:0.01319
[81]#011train-error:0.01319
[82]#011train-error:0.01319
[83]#011train-error:0.01319
[84]#011train-error:0.01319
[85]#011train-error:0.01319
[86]#011train-error:0.01319
[87]#011train-error:0.01319
[88]#011train-error:0.01319
[89]#011train-error:0.01319
[90]#011train-error:0.01319
[91]#011train-error:0.01319
[92]#011train-error:0.01319
[93]#011train-error:0.01319
[94]#011train-error:0.01319
[95]#011train-error:0.01319
[96]#011train-error:0.01319
[97]#011train-error:0.01319
[98]#011train-error:0.01319
[99]#011train-error:0.01319
[100]#011train-error:0.01319
[101]#011train-error:0.01319
[102]#011train-error:0.01319
[103]#011train-error:0.01319
[104]#011train-error:0.01319
[105]#011train-error:0.01319
[106]#011train-error:0.01319
[107]#011train-error:0.01319
[108]#011train-error:0.01319
[109]#011train-error:0.01319
[110]#011train-error:0.01319
[111]#011train-error:0.01319
[112]#011train-error:0.01319

```
[113]#011train-error:0.01319
[114]#011train-error:0.01319
[115]#011train-error:0.01319
[116]#011train-error:0.01319
[117]#011train-error:0.01319
[118]#011train-error:0.01319
[119]#011train-error:0.01319
[120]#011train-error:0.01319
[121]#011train-error:0.01319
[122]#011train-error:0.01319
[123]#011train-error:0.01319
[124]#011train-error:0.01319
[125]#011train-error:0.01319
[126]#011train-error:0.01319
[127]#011train-error:0.01319
[128]#011train-error:0.01319
[129]#011train-error:0.01319
[130]#011train-error:0.01319
[131]#011train-error:0.01319
[132]#011train-error:0.01319
[133]#011train-error:0.01319
[134]#011train-error:0.01319
[135]#011train-error:0.01319
[136]#011train-error:0.01319
[137]#011train-error:0.01319
[138]#011train-error:0.01319
[139]#011train-error:0.01319
[140]#011train-error:0.01319
[141]#011train-error:0.01319
[142]#011train-error:0.01319
[143]#011train-error:0.01319
[144]#011train-error:0.01319
[145]#011train-error:0.01319
[146]#011train-error:0.01319
[147]#011train-error:0.01319
[148]#011train-error:0.01319
[149]#011train-error:0.01319
[150]#011train-error:0.01319
[151]#011train-error:0.01319
[152]#011train-error:0.01319
[153]#011train-error:0.01319
[154]#011train-error:0.01319
[155]#011train-error:0.01319
[156]#011train-error:0.01319
[157]#011train-error:0.01319
[158]#011train-error:0.01319
[159]#011train-error:0.01319
[160]#011train-error:0.01319
```

[161]#011train-error:0.01319
[162]#011train-error:0.01319
[163]#011train-error:0.01319
[164]#011train-error:0.01319
[165]#011train-error:0.01319
[166]#011train-error:0.01319
[167]#011train-error:0.01319
[168]#011train-error:0.01319
[169]#011train-error:0.01319
[170]#011train-error:0.01319
[171]#011train-error:0.01319
[172]#011train-error:0.01319
[173]#011train-error:0.01319
[174]#011train-error:0.01319
[175]#011train-error:0.01099
[176]#011train-error:0.01319
[177]#011train-error:0.01319
[178]#011train-error:0.01319
[179]#011train-error:0.01099
[180]#011train-error:0.01099
[181]#011train-error:0.01099
[182]#011train-error:0.01319
[183]#011train-error:0.01319
[184]#011train-error:0.01319
[185]#011train-error:0.01319
[186]#011train-error:0.01319
[187]#011train-error:0.01319
[188]#011train-error:0.01319
[189]#011train-error:0.01319
[190]#011train-error:0.01319
[191]#011train-error:0.01319
[192]#011train-error:0.01319
[193]#011train-error:0.01319
[194]#011train-error:0.01319
[195]#011train-error:0.01319
[196]#011train-error:0.01319
[197]#011train-error:0.01319
[198]#011train-error:0.01319
[199]#011train-error:0.01319
[200]#011train-error:0.01319
[201]#011train-error:0.01319
[202]#011train-error:0.01319
[203]#011train-error:0.01319
[204]#011train-error:0.01319
[205]#011train-error:0.01319
[206]#011train-error:0.01319
[207]#011train-error:0.01319
[208]#011train-error:0.01319

```
[209]#011train-error:0.01319
[210]#011train-error:0.01319
[211]#011train-error:0.01319
[212]#011train-error:0.01319
[213]#011train-error:0.01319
[214]#011train-error:0.01319
[215]#011train-error:0.01319
[216]#011train-error:0.01319
[217]#011train-error:0.01319
[218]#011train-error:0.01319
[219]#011train-error:0.01099
[220]#011train-error:0.01099
[221]#011train-error:0.01099
[222]#011train-error:0.01099
[223]#011train-error:0.01099
[224]#011train-error:0.01099
[225]#011train-error:0.01099
[226]#011train-error:0.01099
[227]#011train-error:0.01099
[228]#011train-error:0.01099
[229]#011train-error:0.01099
[230]#011train-error:0.01099
[231]#011train-error:0.01099
[232]#011train-error:0.01099
[233]#011train-error:0.01099
[234]#011train-error:0.01099
[235]#011train-error:0.01099
[236]#011train-error:0.01099
[237]#011train-error:0.01099
[238]#011train-error:0.01099
[239]#011train-error:0.01099
[240]#011train-error:0.01099
[241]#011train-error:0.01099
[242]#011train-error:0.01099
[243]#011train-error:0.01099
[244]#011train-error:0.01099
[245]#011train-error:0.01099
[246]#011train-error:0.01099
[247]#011train-error:0.01099
[248]#011train-error:0.01099
[249]#011train-error:0.01099
[250]#011train-error:0.01099
[251]#011train-error:0.01099
[252]#011train-error:0.01099
[253]#011train-error:0.01099
[254]#011train-error:0.01099
[255]#011train-error:0.01099
[256]#011train-error:0.01099
```

[257]#011train-error:0.01099
[258]#011train-error:0.01099
[259]#011train-error:0.01099
[260]#011train-error:0.01099
[261]#011train-error:0.01099
[262]#011train-error:0.01099
[263]#011train-error:0.01099
[264]#011train-error:0.01099
[265]#011train-error:0.01099
[266]#011train-error:0.01099
[267]#011train-error:0.01099
[268]#011train-error:0.01099
[269]#011train-error:0.01099
[270]#011train-error:0.01099
[271]#011train-error:0.01099
[272]#011train-error:0.01099
[273]#011train-error:0.01099
[274]#011train-error:0.01099
[275]#011train-error:0.01099
[276]#011train-error:0.01099
[277]#011train-error:0.01099
[278]#011train-error:0.01099
[279]#011train-error:0.01099
[280]#011train-error:0.01099
[281]#011train-error:0.01099
[282]#011train-error:0.01099
[283]#011train-error:0.01099
[284]#011train-error:0.01099
[285]#011train-error:0.01099
[286]#011train-error:0.01099
[287]#011train-error:0.01099
[288]#011train-error:0.01099
[289]#011train-error:0.01099
[290]#011train-error:0.01099
[291]#011train-error:0.01099
[292]#011train-error:0.01099
[293]#011train-error:0.01099
[294]#011train-error:0.01099
[295]#011train-error:0.01099
[296]#011train-error:0.01099
[297]#011train-error:0.01099
[298]#011train-error:0.01099
[299]#011train-error:0.01099
[300]#011train-error:0.01099
[301]#011train-error:0.01099
[302]#011train-error:0.01099
[303]#011train-error:0.01099
[304]#011train-error:0.01099

[305]#011train-error:0.01099
[306]#011train-error:0.01099
[307]#011train-error:0.01099
[308]#011train-error:0.01099
[309]#011train-error:0.01099
[310]#011train-error:0.01099
[311]#011train-error:0.01099
[312]#011train-error:0.01099
[313]#011train-error:0.01099
[314]#011train-error:0.01099
[315]#011train-error:0.01099
[316]#011train-error:0.01099
[317]#011train-error:0.01099
[318]#011train-error:0.01099
[319]#011train-error:0.01099
[320]#011train-error:0.01099
[321]#011train-error:0.01099
[322]#011train-error:0.01099
[323]#011train-error:0.01099
[324]#011train-error:0.01099
[325]#011train-error:0.01099
[326]#011train-error:0.01099
[327]#011train-error:0.01099
[328]#011train-error:0.01099
[329]#011train-error:0.01099
[330]#011train-error:0.01099
[331]#011train-error:0.01099
[332]#011train-error:0.01099
[333]#011train-error:0.01099
[334]#011train-error:0.01099
[335]#011train-error:0.01099
[336]#011train-error:0.01099
[337]#011train-error:0.01099
[338]#011train-error:0.01099
[339]#011train-error:0.01099
[340]#011train-error:0.01099
[341]#011train-error:0.01099
[342]#011train-error:0.01099
[343]#011train-error:0.01099
[344]#011train-error:0.01099
[345]#011train-error:0.01099
[346]#011train-error:0.01099
[347]#011train-error:0.01099
[348]#011train-error:0.01099
[349]#011train-error:0.01099
[350]#011train-error:0.01099
[351]#011train-error:0.01099
[352]#011train-error:0.01099

[353]#011train-error:0.01099
[354]#011train-error:0.01099
[355]#011train-error:0.01099
[356]#011train-error:0.01099
[357]#011train-error:0.01099
[358]#011train-error:0.01099
[359]#011train-error:0.01099
[360]#011train-error:0.01099
[361]#011train-error:0.01099
[362]#011train-error:0.01099
[363]#011train-error:0.01099
[364]#011train-error:0.01099
[365]#011train-error:0.01099
[366]#011train-error:0.01099
[367]#011train-error:0.01099
[368]#011train-error:0.01099
[369]#011train-error:0.01099
[370]#011train-error:0.01099
[371]#011train-error:0.01099
[372]#011train-error:0.01099
[373]#011train-error:0.01319
[374]#011train-error:0.01319
[375]#011train-error:0.01319
[376]#011train-error:0.01319
[377]#011train-error:0.01319
[378]#011train-error:0.01319
[379]#011train-error:0.01319
[380]#011train-error:0.01319
[381]#011train-error:0.01319
[382]#011train-error:0.01319
[383]#011train-error:0.01319
[384]#011train-error:0.01319
[385]#011train-error:0.01319
[386]#011train-error:0.01319
[387]#011train-error:0.01099
[388]#011train-error:0.01099
[389]#011train-error:0.01099
[390]#011train-error:0.01099
[391]#011train-error:0.01099
[392]#011train-error:0.01099
[393]#011train-error:0.01099
[394]#011train-error:0.01099
[395]#011train-error:0.01099
[396]#011train-error:0.01099
[397]#011train-error:0.01099
[398]#011train-error:0.01099
[399]#011train-error:0.01099
[400]#011train-error:0.01099

[401]#011train-error:0.01099
[402]#011train-error:0.01099
[403]#011train-error:0.01099
[404]#011train-error:0.01099
[405]#011train-error:0.01319
[406]#011train-error:0.01099
[407]#011train-error:0.01099
[408]#011train-error:0.01099
[409]#011train-error:0.01099
[410]#011train-error:0.01099
[411]#011train-error:0.01099
[412]#011train-error:0.01099
[413]#011train-error:0.01099
[414]#011train-error:0.01099
[415]#011train-error:0.01099
[416]#011train-error:0.01099
[417]#011train-error:0.01099
[418]#011train-error:0.01099
[419]#011train-error:0.01099
[420]#011train-error:0.01099
[421]#011train-error:0.01099
[422]#011train-error:0.01099
[423]#011train-error:0.01099
[424]#011train-error:0.01099
[425]#011train-error:0.01099
[426]#011train-error:0.01099
[427]#011train-error:0.01099
[428]#011train-error:0.01099
[429]#011train-error:0.01099
[430]#011train-error:0.01099
[431]#011train-error:0.01099
[432]#011train-error:0.01099
[433]#011train-error:0.01099
[434]#011train-error:0.01099
[435]#011train-error:0.01099
[436]#011train-error:0.01099
[437]#011train-error:0.01099
[438]#011train-error:0.01099
[439]#011train-error:0.01099
[440]#011train-error:0.01099
[441]#011train-error:0.01099
[442]#011train-error:0.01099
[443]#011train-error:0.01099
[444]#011train-error:0.01099
[445]#011train-error:0.01099
[446]#011train-error:0.01099
[447]#011train-error:0.01099
[448]#011train-error:0.01099

[449]#011train-error:0.01099
[450]#011train-error:0.01099
[451]#011train-error:0.01099
[452]#011train-error:0.01099
[453]#011train-error:0.01099
[454]#011train-error:0.01099
[455]#011train-error:0.01099
[456]#011train-error:0.01099
[457]#011train-error:0.01099
[458]#011train-error:0.01099
[459]#011train-error:0.01099
[460]#011train-error:0.01099
[461]#011train-error:0.01099
[462]#011train-error:0.01099
[463]#011train-error:0.01319
[464]#011train-error:0.01319
[465]#011train-error:0.01319
[466]#011train-error:0.01319
[467]#011train-error:0.01319
[468]#011train-error:0.01319
[469]#011train-error:0.01319
[470]#011train-error:0.01319
[471]#011train-error:0.01319
[472]#011train-error:0.01319
[473]#011train-error:0.01319
[474]#011train-error:0.01319
[475]#011train-error:0.01099
[476]#011train-error:0.01099
[477]#011train-error:0.01099
[478]#011train-error:0.01099
[479]#011train-error:0.01099
[480]#011train-error:0.01099
[481]#011train-error:0.01099
[482]#011train-error:0.01099
[483]#011train-error:0.01099
[484]#011train-error:0.01099
[485]#011train-error:0.01099
[486]#011train-error:0.01099
[487]#011train-error:0.01099
[488]#011train-error:0.01099
[489]#011train-error:0.01099
[490]#011train-error:0.01099
[491]#011train-error:0.01099
[492]#011train-error:0.01099
[493]#011train-error:0.01099
[494]#011train-error:0.01099
[495]#011train-error:0.01099
[496]#011train-error:0.01099

[497]#011train-error:0.01099
[498]#011train-error:0.01099
[499]#011train-error:0.01099
[500]#011train-error:0.01319
[501]#011train-error:0.01319
[502]#011train-error:0.01319
[503]#011train-error:0.01319
[504]#011train-error:0.01319
[505]#011train-error:0.01099
[506]#011train-error:0.01099
[507]#011train-error:0.01099
[508]#011train-error:0.01099
[509]#011train-error:0.01099
[510]#011train-error:0.01099
[511]#011train-error:0.01099
[512]#011train-error:0.01099
[513]#011train-error:0.01099
[514]#011train-error:0.01099
[515]#011train-error:0.01099
[516]#011train-error:0.01099
[517]#011train-error:0.01099
[518]#011train-error:0.01099
[519]#011train-error:0.01099
[520]#011train-error:0.01319
[521]#011train-error:0.01319
[522]#011train-error:0.01319
[523]#011train-error:0.01319
[524]#011train-error:0.01099
[525]#011train-error:0.01099
[526]#011train-error:0.01099
[527]#011train-error:0.01099
[528]#011train-error:0.01099
[529]#011train-error:0.01099
[530]#011train-error:0.01099
[531]#011train-error:0.01099
[532]#011train-error:0.01099
[533]#011train-error:0.01099
[534]#011train-error:0.01099
[535]#011train-error:0.01099
[536]#011train-error:0.01099
[537]#011train-error:0.01099
[538]#011train-error:0.01099
[539]#011train-error:0.01099
[540]#011train-error:0.01099
[541]#011train-error:0.01099
[542]#011train-error:0.01099
[543]#011train-error:0.01099
[544]#011train-error:0.01099

[545]#011train-error:0.01099
[546]#011train-error:0.01099
[547]#011train-error:0.01099
[548]#011train-error:0.01099
[549]#011train-error:0.01099
[550]#011train-error:0.01099
[551]#011train-error:0.01099
[552]#011train-error:0.01099
[553]#011train-error:0.01099
[554]#011train-error:0.01099
[555]#011train-error:0.01099
[556]#011train-error:0.01099
[557]#011train-error:0.01099
[558]#011train-error:0.01099
[559]#011train-error:0.01099
[560]#011train-error:0.01099
[561]#011train-error:0.01099
[562]#011train-error:0.01099
[563]#011train-error:0.01099
[564]#011train-error:0.01099
[565]#011train-error:0.01099
[566]#011train-error:0.01099
[567]#011train-error:0.01099
[568]#011train-error:0.01099
[569]#011train-error:0.01099
[570]#011train-error:0.01099
[571]#011train-error:0.01099
[572]#011train-error:0.01099
[573]#011train-error:0.01099
[574]#011train-error:0.01099
[575]#011train-error:0.01099
[576]#011train-error:0.01099
[577]#011train-error:0.01099
[578]#011train-error:0.01099
[579]#011train-error:0.01099
[580]#011train-error:0.01099
[581]#011train-error:0.01099
[582]#011train-error:0.01099
[583]#011train-error:0.01099
[584]#011train-error:0.01099
[585]#011train-error:0.01099
[586]#011train-error:0.01099
[587]#011train-error:0.01099
[588]#011train-error:0.01099
[589]#011train-error:0.01099
[590]#011train-error:0.01099
[591]#011train-error:0.01099
[592]#011train-error:0.01099

[593]#011train-error:0.01099
[594]#011train-error:0.01099
[595]#011train-error:0.01099
[596]#011train-error:0.01099
[597]#011train-error:0.01099
[598]#011train-error:0.01099
[599]#011train-error:0.01099
[600]#011train-error:0.01099
[601]#011train-error:0.01099
[602]#011train-error:0.01099
[603]#011train-error:0.01099
[604]#011train-error:0.01099
[605]#011train-error:0.01099
[606]#011train-error:0.01099
[607]#011train-error:0.01099
[608]#011train-error:0.01099
[609]#011train-error:0.01099
[610]#011train-error:0.01099
[611]#011train-error:0.01099
[612]#011train-error:0.01099
[613]#011train-error:0.01099
[614]#011train-error:0.01099
[615]#011train-error:0.01099
[616]#011train-error:0.01099
[617]#011train-error:0.01099
[618]#011train-error:0.01099
[619]#011train-error:0.01099
[620]#011train-error:0.01099
[621]#011train-error:0.01099
[622]#011train-error:0.01099
[623]#011train-error:0.01099
[624]#011train-error:0.01099
[625]#011train-error:0.01099
[626]#011train-error:0.01099
[627]#011train-error:0.01099
[628]#011train-error:0.01099
[629]#011train-error:0.01099
[630]#011train-error:0.01099
[631]#011train-error:0.01099
[632]#011train-error:0.01099
[633]#011train-error:0.01099
[634]#011train-error:0.01099
[635]#011train-error:0.01099
[636]#011train-error:0.01099
[637]#011train-error:0.01099
[638]#011train-error:0.01099
[639]#011train-error:0.01099
[640]#011train-error:0.01099

[641]#011train-error:0.01099
[642]#011train-error:0.01099
[643]#011train-error:0.01099
[644]#011train-error:0.01099
[645]#011train-error:0.01099
[646]#011train-error:0.01099
[647]#011train-error:0.01099
[648]#011train-error:0.01099
[649]#011train-error:0.01099
[650]#011train-error:0.01099
[651]#011train-error:0.01099
[652]#011train-error:0.01099
[653]#011train-error:0.01099
[654]#011train-error:0.01099
[655]#011train-error:0.01099
[656]#011train-error:0.01099
[657]#011train-error:0.01099
[658]#011train-error:0.01099
[659]#011train-error:0.01099
[660]#011train-error:0.01099
[661]#011train-error:0.01099
[662]#011train-error:0.01099
[663]#011train-error:0.01099
[664]#011train-error:0.01099
[665]#011train-error:0.01099
[666]#011train-error:0.01099
[667]#011train-error:0.01099
[668]#011train-error:0.01099
[669]#011train-error:0.01099
[670]#011train-error:0.01099
[671]#011train-error:0.01099
[672]#011train-error:0.01099
[673]#011train-error:0.01099
[674]#011train-error:0.01099
[675]#011train-error:0.01099
[676]#011train-error:0.01099
[677]#011train-error:0.01099
[678]#011train-error:0.01099
[679]#011train-error:0.01099
[680]#011train-error:0.01099
[681]#011train-error:0.01099
[682]#011train-error:0.01099
[683]#011train-error:0.01099
[684]#011train-error:0.01099
[685]#011train-error:0.01099
[686]#011train-error:0.01099
[687]#011train-error:0.01099
[688]#011train-error:0.01099

[689]#011train-error:0.01099
[690]#011train-error:0.01099
[691]#011train-error:0.01099
[692]#011train-error:0.01099
[693]#011train-error:0.01099
[694]#011train-error:0.01099
[695]#011train-error:0.01099
[696]#011train-error:0.01099
[697]#011train-error:0.01099
[698]#011train-error:0.01099
[699]#011train-error:0.01099
[700]#011train-error:0.01099
[701]#011train-error:0.01099
[702]#011train-error:0.01099
[703]#011train-error:0.01099
[704]#011train-error:0.01099
[705]#011train-error:0.01099
[706]#011train-error:0.01099
[707]#011train-error:0.01099
[708]#011train-error:0.01099
[709]#011train-error:0.01099
[710]#011train-error:0.01099
[711]#011train-error:0.01099
[712]#011train-error:0.01099
[713]#011train-error:0.01099
[714]#011train-error:0.01099
[715]#011train-error:0.01099
[716]#011train-error:0.01099
[717]#011train-error:0.01099
[718]#011train-error:0.01099
[719]#011train-error:0.01099
[720]#011train-error:0.01099
[721]#011train-error:0.01099
[722]#011train-error:0.01099
[723]#011train-error:0.01099
[724]#011train-error:0.01099
[725]#011train-error:0.01099
[726]#011train-error:0.01099
[727]#011train-error:0.01099
[728]#011train-error:0.01099
[729]#011train-error:0.01099
[730]#011train-error:0.01099
[731]#011train-error:0.01099
[732]#011train-error:0.01099
[733]#011train-error:0.01099
[734]#011train-error:0.01099
[735]#011train-error:0.01099
[736]#011train-error:0.01099

[737]#011train-error:0.01099
[738]#011train-error:0.01099
[739]#011train-error:0.01099
[740]#011train-error:0.01099
[741]#011train-error:0.01099
[742]#011train-error:0.01099
[743]#011train-error:0.01099
[744]#011train-error:0.01099
[745]#011train-error:0.01099
[746]#011train-error:0.01099
[747]#011train-error:0.01099
[748]#011train-error:0.01099
[749]#011train-error:0.01099
[750]#011train-error:0.01099
[751]#011train-error:0.01099
[752]#011train-error:0.01099
[753]#011train-error:0.01099
[754]#011train-error:0.01099
[755]#011train-error:0.01099
[756]#011train-error:0.01099
[757]#011train-error:0.01099
[758]#011train-error:0.01099
[759]#011train-error:0.01099
[760]#011train-error:0.01099
[761]#011train-error:0.01099
[762]#011train-error:0.01099
[763]#011train-error:0.01099
[764]#011train-error:0.01099
[765]#011train-error:0.01099
[766]#011train-error:0.01099
[767]#011train-error:0.01099
[768]#011train-error:0.01099
[769]#011train-error:0.01099
[770]#011train-error:0.01099
[771]#011train-error:0.01099
[772]#011train-error:0.01099
[773]#011train-error:0.01099
[774]#011train-error:0.01099
[775]#011train-error:0.01099
[776]#011train-error:0.01099
[777]#011train-error:0.01099
[778]#011train-error:0.01099
[779]#011train-error:0.01099
[780]#011train-error:0.01099
[781]#011train-error:0.01099
[782]#011train-error:0.01099
[783]#011train-error:0.01099
[784]#011train-error:0.01099

[785]#011train-error:0.01099
[786]#011train-error:0.01099
[787]#011train-error:0.01099
[788]#011train-error:0.01099
[789]#011train-error:0.01099
[790]#011train-error:0.01099
[791]#011train-error:0.01099
[792]#011train-error:0.01099
[793]#011train-error:0.01099
[794]#011train-error:0.01099
[795]#011train-error:0.01099
[796]#011train-error:0.01099
[797]#011train-error:0.01099
[798]#011train-error:0.01099
[799]#011train-error:0.01099
[800]#011train-error:0.01099
[801]#011train-error:0.01099
[802]#011train-error:0.01099
[803]#011train-error:0.01099
[804]#011train-error:0.01099
[805]#011train-error:0.01099
[806]#011train-error:0.01099
[807]#011train-error:0.01099
[808]#011train-error:0.01099
[809]#011train-error:0.01099
[810]#011train-error:0.01099
[811]#011train-error:0.01099
[812]#011train-error:0.01099
[813]#011train-error:0.01099
[814]#011train-error:0.01099
[815]#011train-error:0.01099
[816]#011train-error:0.01099
[817]#011train-error:0.01099
[818]#011train-error:0.01099
[819]#011train-error:0.01099
[820]#011train-error:0.01099
[821]#011train-error:0.01099
[822]#011train-error:0.01099
[823]#011train-error:0.01099
[824]#011train-error:0.01099
[825]#011train-error:0.01099
[826]#011train-error:0.01099
[827]#011train-error:0.01099
[828]#011train-error:0.01099
[829]#011train-error:0.01099
[830]#011train-error:0.01099
[831]#011train-error:0.01099
[832]#011train-error:0.01099

[833]#011train-error:0.01099
[834]#011train-error:0.01099
[835]#011train-error:0.01099
[836]#011train-error:0.01099
[837]#011train-error:0.01099
[838]#011train-error:0.01099
[839]#011train-error:0.01099
[840]#011train-error:0.01099
[841]#011train-error:0.01099
[842]#011train-error:0.01099
[843]#011train-error:0.01099
[844]#011train-error:0.01099
[845]#011train-error:0.01099
[846]#011train-error:0.01099
[847]#011train-error:0.01099
[848]#011train-error:0.01099
[849]#011train-error:0.01099
[850]#011train-error:0.01099
[851]#011train-error:0.01099
[852]#011train-error:0.01099
[853]#011train-error:0.01099
[854]#011train-error:0.01099
[855]#011train-error:0.01099
[856]#011train-error:0.01099
[857]#011train-error:0.01099
[858]#011train-error:0.01099
[859]#011train-error:0.01099
[860]#011train-error:0.01099
[861]#011train-error:0.01099
[862]#011train-error:0.01099
[863]#011train-error:0.01099
[864]#011train-error:0.01099
[865]#011train-error:0.01099
[866]#011train-error:0.01099
[867]#011train-error:0.01099
[868]#011train-error:0.01099
[869]#011train-error:0.01099
[870]#011train-error:0.01099
[871]#011train-error:0.01099
[872]#011train-error:0.01099
[873]#011train-error:0.01099
[874]#011train-error:0.01099
[875]#011train-error:0.01099
[876]#011train-error:0.01099
[877]#011train-error:0.01099
[878]#011train-error:0.01099
[879]#011train-error:0.01099
[880]#011train-error:0.01099

[881]#011train-error:0.01099
[882]#011train-error:0.01099
[883]#011train-error:0.01099
[884]#011train-error:0.01099
[885]#011train-error:0.01099
[886]#011train-error:0.01099
[887]#011train-error:0.01099
[888]#011train-error:0.01099
[889]#011train-error:0.01099
[890]#011train-error:0.01099
[891]#011train-error:0.01099
[892]#011train-error:0.01099
[893]#011train-error:0.01099
[894]#011train-error:0.01099
[895]#011train-error:0.01099
[896]#011train-error:0.01099
[897]#011train-error:0.01099
[898]#011train-error:0.01099
[899]#011train-error:0.01099
[900]#011train-error:0.01099
[901]#011train-error:0.01099
[902]#011train-error:0.01099
[903]#011train-error:0.01099
[904]#011train-error:0.01099
[905]#011train-error:0.01099
[906]#011train-error:0.01099
[907]#011train-error:0.01099
[908]#011train-error:0.01099
[909]#011train-error:0.01099
[910]#011train-error:0.01099
[911]#011train-error:0.01099
[912]#011train-error:0.01099
[913]#011train-error:0.01099
[914]#011train-error:0.01099
[915]#011train-error:0.01099
[916]#011train-error:0.01099
[917]#011train-error:0.01099
[918]#011train-error:0.01099
[919]#011train-error:0.01099
[920]#011train-error:0.01099
[921]#011train-error:0.01099
[922]#011train-error:0.01099
[923]#011train-error:0.01099
[924]#011train-error:0.01099
[925]#011train-error:0.01099
[926]#011train-error:0.01099
[927]#011train-error:0.01099
[928]#011train-error:0.01099

[929]#011train-error:0.01099
[930]#011train-error:0.01099
[931]#011train-error:0.01099
[932]#011train-error:0.01099
[933]#011train-error:0.01099
[934]#011train-error:0.01099
[935]#011train-error:0.01099
[936]#011train-error:0.01099
[937]#011train-error:0.01099
[938]#011train-error:0.01099
[939]#011train-error:0.01099
[940]#011train-error:0.01099
[941]#011train-error:0.01099
[942]#011train-error:0.01099
[943]#011train-error:0.01099
[944]#011train-error:0.01099
[945]#011train-error:0.01099
[946]#011train-error:0.01099
[947]#011train-error:0.01099
[948]#011train-error:0.01099
[949]#011train-error:0.01099
[950]#011train-error:0.01099
[951]#011train-error:0.01099
[952]#011train-error:0.01099
[953]#011train-error:0.01099
[954]#011train-error:0.01099
[955]#011train-error:0.01099
[956]#011train-error:0.01099
[957]#011train-error:0.01099
[958]#011train-error:0.01099
[959]#011train-error:0.01099
[960]#011train-error:0.01099
[961]#011train-error:0.01099
[962]#011train-error:0.01099
[963]#011train-error:0.01099
[964]#011train-error:0.01099
[965]#011train-error:0.01099
[966]#011train-error:0.01099
[967]#011train-error:0.01099
[968]#011train-error:0.01099
[969]#011train-error:0.01099
[970]#011train-error:0.01099
[971]#011train-error:0.01099
[972]#011train-error:0.01099
[973]#011train-error:0.01099
[974]#011train-error:0.01099
[975]#011train-error:0.01099
[976]#011train-error:0.01099

```
[977]#011train-error:0.01099
[978]#011train-error:0.01099
[979]#011train-error:0.01099
[980]#011train-error:0.01099
[981]#011train-error:0.01099
[982]#011train-error:0.01099
[983]#011train-error:0.01099
[984]#011train-error:0.01099
[985]#011train-error:0.01099
[986]#011train-error:0.01099
[987]#011train-error:0.01099
[988]#011train-error:0.01099
[989]#011train-error:0.01099
[990]#011train-error:0.01099
[991]#011train-error:0.01099
[992]#011train-error:0.01099
[993]#011train-error:0.01099
[994]#011train-error:0.01099
[995]#011train-error:0.01099
[996]#011train-error:0.01099
[997]#011train-error:0.01099
[998]#011train-error:0.01099
[999]#011train-error:0.01099

2020-12-04 19:37:11 Uploading - Uploading generated training model
2020-12-04 19:37:11 Completed - Training job completed
Training seconds: 71
Billable seconds: 71
```

[71]:
```
# Creating a model endpoint for testing
xgb_predictor = xgb.deploy(initial_instance_count=1, serializer =␣
 ↪CSVSerializer(), instance_type='ml.m4.xlarge')
```

```
----------------!
```

[72]:
```
test_data_array = test_data.drop('diagnosis', axis = 1).values
```

[73]:
```
# Predictions are output as a string of probabilities separated by ','.
# That is converted to an array.
predictions = xgb_predictor.predict(test_data_array).decode('utf-8')
predictions_array = np.fromstring(predictions[1:], sep=',')
```

[75]:
```
# Probabilities to class labels
predictions_array = np.where(predictions_array > 0.5, 1, 0)
predictions_array
```

[75]:
```
array([0, 0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1,
       0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0,
       1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0,
       0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1,
       1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0,
       0, 0, 1, 0])
```

[85]:
```
from sklearn.metrics import accuracy_score, roc_curve, roc_auc_score
accuracy_score(test_data['diagnosis'], predictions_array)
```

[85]:
```
0.9473684210526315
```

[86]:
```
# Calculating the roc_curve parameters - TPR, FPR, Area Under the Curve (AUC)
ns_labels = [0 for _ in range(len(test_data['diagnosis']))]
ns_fpr, ns_tpr, _ = roc_curve(test_data['diagnosis'], ns_labels)
xgb_fpr, xgb_tpr, _ = roc_curve(test_data['diagnosis'], predictions_array)
AUC = roc_auc_score(test_data['diagnosis'], predictions_array)
```

[89]:
```
# Analysis of ROC curve
plt.plot(ns_fpr, ns_tpr, linestyle = '--', label = "No-skill Classifier")
plt.plot(fpr, tpr, marker = '*', label = "XGBoost Classifier")
plt.xlabel('False Positive Rate')
plt.xlabel('True Positive Rate')
plt.text(0.3, 0.6, "AUC = {}".format(np.round(AUC,3)))
plt.legend()
```

[89]:
```
<matplotlib.legend.Legend at 0x7fcb17dfbba8>
```