3.0-SMOTE

December 31, 2023

0.1 SMOTE(Synthetic Minority Oversampling Technique)

SMOTE (Synthetic Minority Over-sampling Technique) is a technique used in machine learning to address imbalanced datasets where the minority class has significantly fewer instances than the majority class. SMOTE involves generating synthetic instances of the minority class by interpolating between existing instances.

[ ]:
```
from sklearn.datasets import make_classification
```

[ ]:
```
X,y=make_classification(n_samples=1000,n_redundant=0,n_features=2,n_clusters_per_class=1,
                    weights=[0.90],random_state=12)
```

[ ]:
```
X.shape
```

[ ]:
```
(1000, 2)
```

[ ]:
```
len(y[y==1])
```

[ ]:
```
100
```

[ ]:
```

```

[ ]:
```

```

[ ]:
```

```

[ ]:
```
import pandas as pd
df1=pd.DataFrame(X,columns=['f1','f2'])
df2=pd.DataFrame(y,columns=['target'])
final_df=pd.concat([df1,df2],axis=1)
final_df.head()
```

[ ]:

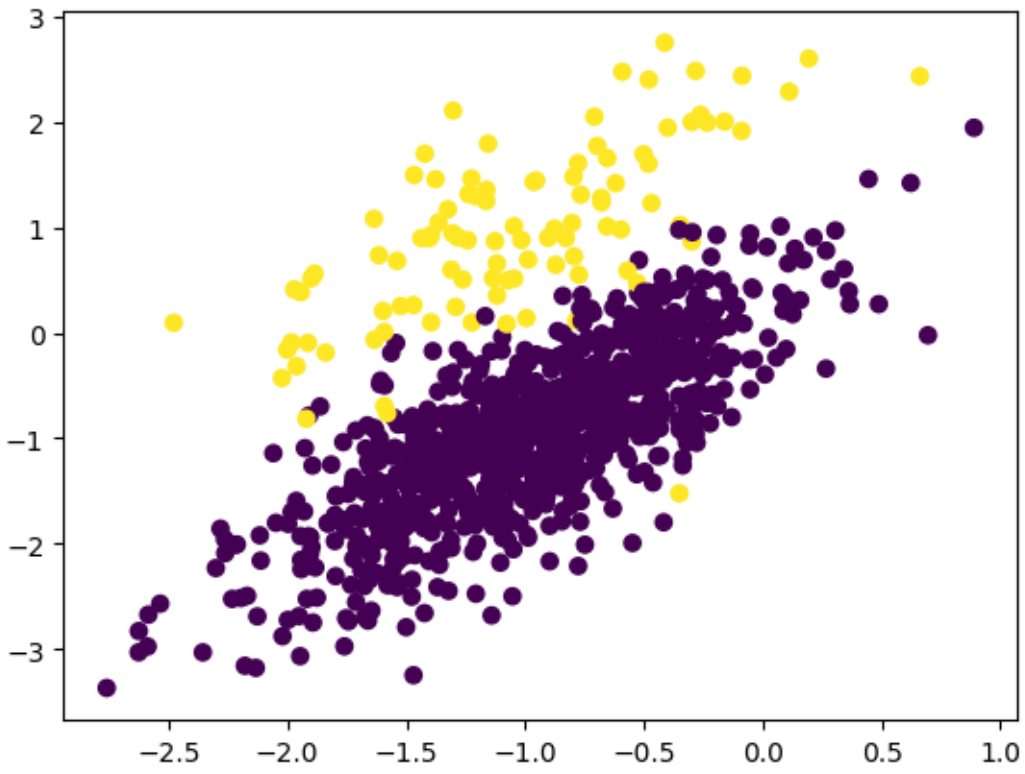
	f1	f2	target
0	-0.762898	-0.706808	0
1	-1.075436	-1.051162	0
2	-0.610115	-0.909802	0
3	-2.023284	-0.428945	1
4	-0.812921	-1.316206	0

[ ]:
```
final_df['target'].value_counts()
```

[ ]:
```
0    900
1    100
Name: target, dtype: int64
```

[ ]:
```
import matplotlib.pyplot as plt
plt.scatter(final_df['f1'],final_df['f2'],c=final_df['target'])
```

[ ]:
```
<matplotlib.collections.PathCollection at 0x7a8e73a0c280>
```



[ ]:
```
!pip install imblearn
```

```
Collecting imblearn
  Downloading imblearn-0.0-py2.py3-none-any.whl (1.9 kB)
Requirement already satisfied: imbalanced-learn in
/usr/local/lib/python3.10/dist-packages (from imblearn) (0.10.1)
Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.10/dist-
packages (from imbalanced-learn->imblearn) (1.23.5)
Requirement already satisfied: scipy>=1.3.2 in /usr/local/lib/python3.10/dist-
packages (from imbalanced-learn->imblearn) (1.10.1)
Requirement already satisfied: scikit-learn>=1.0.2 in
```

```
/usr/local/lib/python3.10/dist-packages (from imbalanced-learn->imblearn)
(1.2.2)
Requirement already satisfied: joblib>=1.1.1 in /usr/local/lib/python3.10/dist-
packages (from imbalanced-learn->imblearn) (1.3.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/usr/local/lib/python3.10/dist-packages (from imbalanced-learn->imblearn)
(3.2.0)
Installing collected packages: imblearn
Successfully installed imblearn-0.0
```

[ ]:
```
from imblearn.over_sampling import SMOTE
```

[ ]:
```
## transform the dataset
oversample=SMOTE()
X,y=oversample.fit_resample(final_df[['f1','f2']],final_df['target'])
```

[ ]:
```
X.shape
```

[ ]:
```
(1800, 2)
```

[ ]:
```
y.shape
```

[ ]:
```
(1800,)
```

[ ]:
```
len(y[y==0])
```

[ ]:
```
900
```

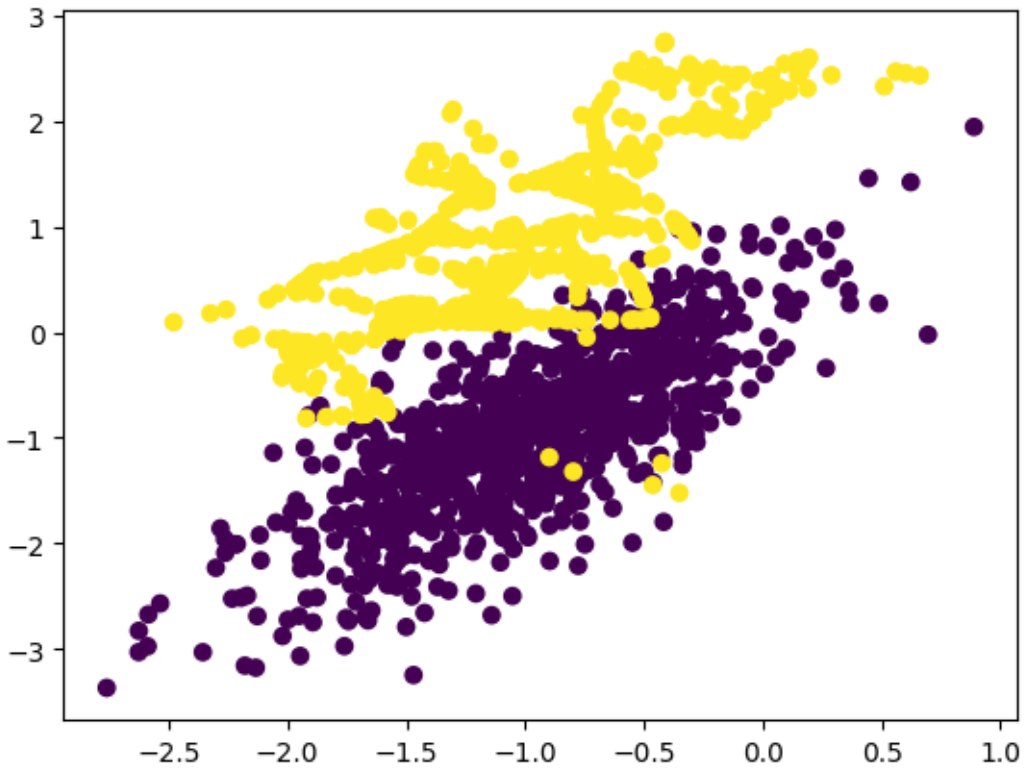
[ ]:
```
len(y[y==1])
```

[ ]:
```
900
```

[ ]:
```
df1=pd.DataFrame(X,columns=['f1','f2'])
df2=pd.DataFrame(y,columns=['target'])
oversample_df=pd.concat([df1,df2],axis=1)
```

[ ]:
```
plt.scatter(oversample_df['f1'],oversample_df['f2'],c=oversample_df['target'])
```

[ ]:
```
<matplotlib.collections.PathCollection at 0x7a8e73903d90>
```


3
2
1
0
−1
−2
−3
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: