

DecisionTreeClassifier

January 13, 2022

0.1 DecisionTreeClassifier

Import libraries

```
[60]: import pandas as pd
      from sklearn.tree import DecisionTreeClassifier
      from sklearn.model_selection import train_test_split
      from sklearn.metrics import accuracy_score
      import matplotlib.pyplot as plt
```

Load data

```
[61]: music_data = pd.read_csv("../Music/music.csv")
      music_data
```

```
[61]:
```

	age	gender	genre
0	20	1	HipHop
1	23	1	HipHop
2	25	1	HipHop
3	26	1	Jazz
4	29	1	Jazz
5	30	1	Jazz
6	31	1	Classical
7	33	1	Classical
8	37	1	Classical
9	20	0	Dance
10	21	0	Dance
11	25	0	Dance
12	26	0	Acoustic
13	27	0	Acoustic
14	30	0	Acoustic
15	31	0	Classical
16	34	0	Classical
17	35	0	Classical

Get dummies to go from categorical to numeric

```
[62]: music_data_dummies = pd.get_dummies(music_data, columns= ['genre'],
↳drop_first=True)
music_data_dummies
```

```
[62]:
```

	age	gender	genre_Classical	genre_Dance	genre_HipHop	genre_Jazz
0	20	1	0	0	1	0
1	23	1	0	0	1	0
2	25	1	0	0	1	0
3	26	1	0	0	0	1
4	29	1	0	0	0	1
5	30	1	0	0	0	1
6	31	1	1	0	0	0
7	33	1	1	0	0	0
8	37	1	1	0	0	0
9	20	0	0	1	0	0
10	21	0	0	1	0	0
11	25	0	0	1	0	0
12	26	0	0	0	0	0
13	27	0	0	0	0	0
14	30	0	0	0	0	0
15	31	0	1	0	0	0
16	34	0	1	0	0	0
17	35	0	1	0	0	0

Check for missing values

```
[63]: music_data_dummies.isnull().sum()
```

```
[63]: age                0
gender                0
genre_Classical      0
genre_Dance          0
genre_HipHop         0
genre_Jazz           0
dtype: int64
```

Training a model

```
[64]: X = music_data_dummies.
↳drop(columns=['genre_Classical','genre_Dance','genre_HipHop','genre_Jazz'])
y =
↳music_data_dummies[['genre_Classical','genre_Dance','genre_HipHop','genre_Jazz']]

X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.2)

model = DecisionTreeClassifier()
```

```
history = model.fit(X_train,y_train)
history
predictions = model.predict(X_test)
score = accuracy_score(y_test,predictions)
score
```

[64]: 1.0