implementation-using-iris-dataset

July 23, 2023

[1]:
```
import numpy as np
import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import load_iris
```

[2]:
```
# Load the Iris dataset
iris = load_iris()
data = iris.data
target = iris.target
feature_names = iris.feature_names
target_names = iris.target_names
```

[3]:
```
# Convert data and target to a pandas DataFrame
df = pd.DataFrame(data, columns=feature_names)
df['target'] = target
```

[4]:
```
# Function to calculate entropy
def entropy(target):
    _, counts = np.unique(target, return_counts=True)
    probabilities = counts / len(target)
    entropy = sum([-p * np.log2(p) for p in probabilities])
    return entropy
```

[5]:
```
# Function to calculate information gain
def information_gain(data, feature, target):
    original_entropy = entropy(target)
    unique_values = np.unique(data[feature])
    weighted_entropy = sum([
        (np.sum(data[feature] == value) / len(data)) *␣
 ↪entropy(target[data[feature] == value])
        for value in unique_values
    ])
    gain = original_entropy - weighted_entropy
    return gain
```

[6]:
```
# Function to build the decision tree
def build_decision_tree(data, target, features, level=0):
```

```
    # Get unique target values
    unique_targets = np.unique(target)

    # Print the current level
    print('Level', level)

    # Print the count of each target value
    for target_value in unique_targets:
        count = np.sum(target == target_value)
        print('Count of', target_names[target_value], '=', count)

    # Print the current entropy
    current_entropy = entropy(target)
    print('Current Entropy is =', current_entropy)

    # Check if all instances belong to the same class or no features are left
    if len(unique_targets) == 1 or len(features) == 0:
        print('Reached leaf Node')
        return

    # Calculate the information gain for each feature
    gains = [information_gain(data, feature, target) for feature in features]

    # Get the index of the feature with the maximum gain
    best_feature_index = np.argmax(gains)
    best_feature = features[best_feature_index]

    # Print the splitting feature and its gain ratio
    print('Splitting on feature', best_feature, 'with gain ratio',␣
↪gains[best_feature_index])

    # Get unique values of the selected feature
    unique_values = np.unique(data[best_feature])

    # Remove the selected feature from the feature list
    features = features.drop(best_feature)

    # Recursively build the decision tree for each unique value
    for value in unique_values:
        print('\n')
        print('Level', level + 1)

        # Create subsets of the data and target based on the selected feature␣
↪value
        subset_data = data[data[best_feature] == value]
        subset_target = target[data[best_feature] == value]
```

```
        # Check if the subset is empty
        if len(subset_data) == 0:
            print('Count of', target_names[unique_targets[0]], '=', np.
 ↪sum(target == unique_targets[0]))
            print('Current Entropy is =', entropy(target))
            print('Reached leaf Node')
            continue

        # Print the count of each target value in the subset
        for target_value in unique_targets:
            count = np.sum(subset_target == target_value)
            print('Count of', target_names[target_value], '=', count)

        # Print the current entropy of the subset
        current_entropy = entropy(subset_target)
        print('Current Entropy is =', current_entropy)

        # Recursively build the decision tree for the subset
        build_decision_tree(subset_data, subset_target, features, level + 1)
```

[7]:
```
# Prepare the data
features = df.columns[:-1]
```

[8]:
```
# Build the decision tree
build_decision_tree(df, target, features)
```

```
Level 0
Count of setosa = 50
Count of versicolor = 50
Count of virginica = 50
Current Entropy is = 1.584962500721156
Splitting on feature petal length (cm) with gain ratio 1.4463165236458


Level 1
Count of setosa = 1
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 1
```

```
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 2
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 7
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 7
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 13
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 13
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 13
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
```

```
Count of setosa = 13
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 7
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 7
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 4
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 4
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 2
Count of versicolor = 0
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of setosa = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node
```

```
Level 1
Count of setosa = 0
Count of versicolor = 2
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 2
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 1
```

```
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 3
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 3
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 5
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 5
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 3
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 3
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 4
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 4
```

```
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 2
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 4
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 4
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 7
Count of virginica = 1
Current Entropy is = 0.5435644431995964
Level 1
Count of versicolor = 7
Count of virginica = 1
Current Entropy is = 0.5435644431995964
Splitting on feature sepal length (cm) with gain ratio 0.5435644431995964


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node
```

```
Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 2
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 2
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node
```

```
Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 3
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 3
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 5
Count of virginica = 0
Current Entropy is = 0.0
Level 1
Count of versicolor = 5
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 2
Count of virginica = 2
Current Entropy is = 1.0
Level 1
Count of versicolor = 2
Count of virginica = 2
Current Entropy is = 1.0
Splitting on feature sepal length (cm) with gain ratio 1.0


Level 2
Count of versicolor = 1
```

```
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 2
Count of virginica = 3
Current Entropy is = 0.9709505944546686
Level 1
Count of versicolor = 2
Count of virginica = 3
Current Entropy is = 0.9709505944546686
Splitting on feature sepal width (cm) with gain ratio 0.9709505944546686
```

```
Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
```

```
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 1
Count of virginica = 3
Current Entropy is = 0.8112781244591328
Level 1
Count of versicolor = 1
Count of virginica = 3
Current Entropy is = 0.8112781244591328
Splitting on feature sepal length (cm) with gain ratio 0.8112781244591328


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
```

```
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 1
Count of virginica = 7
Current Entropy is = 0.5435644431995964
Level 1
Count of versicolor = 1
Count of virginica = 7
Current Entropy is = 0.5435644431995964
Splitting on feature sepal length (cm) with gain ratio 0.5435644431995964


Level 2
Count of versicolor = 0
Count of virginica = 3
Current Entropy is = 0.0
Level 2
Count of virginica = 3
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 1
Count of virginica = 0
Current Entropy is = 0.0
Level 2
Count of versicolor = 1
Current Entropy is = 0.0
Reached leaf Node
```

```
Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 2
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 2
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 2
Current Entropy is = 0.0
Level 1
Count of virginica = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 2
Current Entropy is = 0.0
```

```
Level 1
Count of virginica = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 2
Current Entropy is = 0.0
Level 1
Count of virginica = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 3
Current Entropy is = 0.0
Level 1
Count of virginica = 3
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 6
Current Entropy is = 0.0
Level 1
Count of virginica = 6
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 3
Current Entropy is = 0.0
Level 1
Count of virginica = 3
Current Entropy is = 0.0
Reached leaf Node
```

```
Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 3
Current Entropy is = 0.0
Level 1
Count of virginica = 3
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 2
Current Entropy is = 0.0
Level 1
Count of virginica = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 2
Current Entropy is = 0.0
Level 1
Count of virginica = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 3
Current Entropy is = 0.0
Level 1
Count of virginica = 3
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
```

```
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 1
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 1
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 1
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 2
Current Entropy is = 0.0
Level 1
Count of virginica = 2
Current Entropy is = 0.0
Reached leaf Node


Level 1
Count of setosa = 0
Count of versicolor = 0
Count of virginica = 1
Current Entropy is = 0.0
Level 1
```

```
Count of virginica = 1
Current Entropy is = 0.0
Reached leaf Node
```

0.0.1 Python code for document containing the OR Tree formed using the python libraries- pydotplus and graphviz, Format: .pdf

[10]:
```
import pydotplus
from sklearn.tree import export_graphviz
```

[11]:
```
# Define the OR dataset
X = [[0, 0], [0, 1], [1, 0], [1, 1]]
Y = [0, 1, 1, 1]
```

[12]:
```
# Create a decision tree classifier
clf = DecisionTreeClassifier()
clf.fit(X, Y)
```

[12]:
```
DecisionTreeClassifier()
```

[13]:
```
# Generate the dot file
dot_data = export_graphviz(clf, out_file=None, feature_names=['X1', 'X2'],␣
 ↪class_names=['False', 'True'], filled=True, rounded=True,␣
 ↪special_characters=True)
```

[15]:
```
# Create the graph from the dot data
graph = pydotplus.graph_from_dot_data(dot_data)
```

[16]:
```
# Write the graph to a file
graph.write_pdf("or_tree.pdf")
```

[16]:
```
True
```

0.0.2 Python code for a document containing the Iris Dataset Tree formed using the python libraries- pydotplus and graphviz, Format: .pdf

[23]:
```
# Create a decision tree classifier
clf = DecisionTreeClassifier()
clf.fit(iris.data, iris.target)
```

[23]:
```
DecisionTreeClassifier()
```

[25]:
```
# Create the graph from the dot data
dot_data = export_graphviz(clf, out_file = None,
                           feature_names = iris.feature_names,
                           class_names = iris.target_names)
graph = pydotplus.graph_from_dot_data(dot_data)
```

[26]:
```
# Write the graph to a file
graph.write_pdf("iris_tree.pdf")
```

[26]: True