Week-5

- Decision tree ID3

```
import numpy as np
import pandas as pd
df = pd.read_csv('url')
df.head()
df.info()
df.describe()
def find_entropy(df):
    target = df.keys()[-1]
    entropy = 0
    values = df[target].unique()
    for value in values:
        fraction = df[target].value_counts()[value]/
                   len(df[target])
        entropy += fraction * np.log2(fraction)
    return entropy
def buildtree(df, tree=None):
    target = df.keys()[-1]
    node = find_winner(df)
    att = np.unique(df[node])
    if tree is None:
        tree = {}
    tree[node] = {}
    for value in att:
        sub = get_subtable(df, node, value)
    clvalue, counts = np.unique(subtable[target],
                                return_counts = True)
    if len(counts) == 1:
        tree[node][value] = value[0]
    else:
        tree[node][value] = buildtree(subtable)
```

```
return tree
tree = buildtree(df)
import pprint
pprint.print(tree)
```

- Decision tree (sklearn):

```
import pandas as pd
import numpy as np
import sklearn.model_selection import DecisionTree
from sklearn.tree import plot_tree
df = pd.read_csv('iris')
    df.head()
    df.info()
    df.isnull().sum()
    cols = df.columns[0:-1]
    for i in cols:
        sns.boxplot(y=df[i])
        plt.show()
    x = df.drop('species', axis=1)
    y = df['species']
x_train, x_test, y_train, y_test = train_test_
                    split(x, y, test_size=0.3)
dt = DecisionTreeClassifier(max_depth=3)
dt.fit(x, y)
y_pred_train = dt.predict(x_train)
y_pred = dt.predict(x_test)

accuracy_score(y_pred, y_test)
```

25 09.05.24