Report - Assignment 2
Machine Learning (CSE343), Monsoon 2021

- Abhimanyu Gupta
2019226

Q1
Explanations:

a.

Entropy performs better than Gini (by a margin of as close to ~2%).

b.

As the depth of the decision tree increases, accuracy increases as well. As we are fitting the model on the training set, its accuracy is more than the testing set.

c.

Because of the max-depth set to 3, the accuracy of this model is very less as compared to a single decision tree in previous parts.

d.

Using multiple decision stumps and taking majority votes, accuracy increases several folds as compared to a single decision tree.
Random Forest with atleast 25 stumps and depth greater than 20 performs the best.
Specifically with 100 stumps and 20 max-depth best performance is observed.

e.

Mine Implementation of Random Forest performs better than sklearn's Adaboost.

Plots:

Outputs:

```
Part A :
entropy gives better accuray on the testing set, with a performance of 0.8279847908745247
```

```
Part B :
At depth = 30 maximum performance is observed. Training Score: 1.0 Testing Score: 0.8263117870722434
```

```
Part C :
For Decision Stumps = 100 , Maximum Depth = 3 , Used training set ratio = 0.5 :-
Training Accuracy = 0.3441126613639327 Testing Accuracy = 0.3379467680608365
```

```
Part D :
For Decision Stumps = 10 , Maximum Depth = 4 , Used training set ratio = 0.5 :-
Training Accuracy = 0.40200808449602293 Validation Accuracy = 0.39525330899132816 Testing Accuracy = 0.39756653992395435
For Decision Stumps = 10 , Maximum Depth = 8 , Used training set ratio = 0.5 :-
Training Accuracy = 0.6282109792671795 Validation Accuracy = 0.5982047771185152 Testing Accuracy = 0.6069961977186312
For Decision Stumps = 10 , Maximum Depth = 10 , Used training set ratio = 0.5 :-
Training Accuracy = 0.7813600208632155 Validation Accuracy = 0.7211319032405294 Testing Accuracy = 0.7321673003802281
For Decision Stumps = 10 , Maximum Depth = 15 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9754531229625766 Validation Accuracy = 0.8629240833713677 Testing Accuracy = 0.8736121673003803
For Decision Stumps = 10 , Maximum Depth = 20 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9870908853827096 Validation Accuracy = 0.8811805872508748 Testing Accuracy = 0.8768060836501901
For Decision Stumps = 10 , Maximum Depth = 30 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9881340461598644 Validation Accuracy = 0.8753993610223643 Testing Accuracy = 0.87893536121673
For Decision Stumps = 25 , Maximum Depth = 4 , Used training set ratio = 0.5 :-
Training Accuracy = 0.40380101708175775 Validation Accuracy = 0.3943404837973528 Testing Accuracy = 0.39726235741444865
For Decision Stumps = 25 , Maximum Depth = 8 , Used training set ratio = 0.5 :-
Training Accuracy = 0.6458469161559526 Validation Accuracy = 0.6202647193062528 Testing Accuracy = 0.6209885931558935
For Decision Stumps = 25 , Maximum Depth = 10 , Used training set ratio = 0.5 :-
Training Accuracy = 0.8050267309949146 Validation Accuracy = 0.7529286474973376 Testing Accuracy = 0.7473764258555133
For Decision Stumps = 25 , Maximum Depth = 15 , Used training set ratio = 0.5 :-
Training Accuracy = 0.988394836354153 Validation Accuracy = 0.8944165525635174 Testing Accuracy = 0.8923193916349815
For Decision Stumps = 25 , Maximum Depth = 20 , Used training set ratio = 0.5 :-
```

```
For Decision Stumps = 25 , Maximum Depth = 20 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9981418698656931 Validation Accuracy = 0.9064354176175262 Testing Accuracy = 0.9034220532319391
For Decision Stumps = 25 , Maximum Depth = 30 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9971639066371104 Validation Accuracy = 0.9090217556671231 Testing Accuracy = 0.9015969581749049
For Decision Stumps = 50 , Maximum Depth = 4 , Used training set ratio = 0.5 :-
Training Accuracy = 0.41543877950189073 Validation Accuracy = 0.4031644606724479 Testing Accuracy = 0.4073003802281368
For Decision Stumps = 50 , Maximum Depth = 8 , Used training set ratio = 0.5 :-
Training Accuracy = 0.6450971443473725 Validation Accuracy = 0.6155484558040468 Testing Accuracy = 0.6135361216730038
For Decision Stumps = 50 , Maximum Depth = 10 , Used training set ratio = 0.5 :-
Training Accuracy = 0.8110901030121267 Validation Accuracy = 0.7529286474973376 Testing Accuracy = 0.7546768060836502
For Decision Stumps = 50 , Maximum Depth = 15 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9934476463684966 Validation Accuracy = 0.9030883919062833 Testing Accuracy = 0.900532319391635
For Decision Stumps = 50 , Maximum Depth = 20 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9992176294171339 Validation Accuracy = 0.9119123687813784 Testing Accuracy = 0.9133079847908745
For Decision Stumps = 50 , Maximum Depth = 30 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9989568392228452 Validation Accuracy = 0.910999543587403 Testing Accuracy = 0.9072243346007605
For Decision Stumps = 100 , Maximum Depth = 4 , Used training set ratio = 0.5 :-
Training Accuracy = 0.4029534489503195 Validation Accuracy = 0.3937319336680359 Testing Accuracy = 0.39543726235741444
For Decision Stumps = 100 , Maximum Depth = 8 , Used training set ratio = 0.5 :-
Training Accuracy = 0.6452927369930891 Validation Accuracy = 0.6146356306100715 Testing Accuracy = 0.6097338403041825
For Decision Stumps = 100 , Maximum Depth = 10 , Used training set ratio = 0.5 :-
Training Accuracy = 0.8138609988264441 Validation Accuracy = 0.7574927734672143 Testing Accuracy = 0.7606083650190114
For Decision Stumps = 100 , Maximum Depth = 15 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9951101838570869 Validation Accuracy = 0.9046097672295755 Testing Accuracy = 0.9047908745247148
For Decision Stumps = 100 , Maximum Depth = 20 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9996740122571391 Validation Accuracy = 0.9172371824129013 Testing Accuracy = 0.9166539923954372
For Decision Stumps = 100 , Maximum Depth = 30 , Used training set ratio = 0.5 :-
Training Accuracy = 0.9997392098057113 Validation Accuracy = 0.9161722196865967 Testing Accuracy = 0.912851711026616
```

```
Part E :
For Number of Estimator = 4 , Criterion = entropy , Maximum Depth = 30 :-
Training Accuracy = 1.0 Validation Accuracy = 0.8256503879507074 Testing Accuracy = 0.8273764258555133
For Number of Estimator = 8 , Criterion = entropy , Maximum Depth = 30 :-
Training Accuracy = 1.0 Validation Accuracy = 0.8270196257416704 Testing Accuracy = 0.8304182509505703
For Number of Estimator = 10 , Criterion = entropy , Maximum Depth = 30 :-
Training Accuracy = 1.0 Validation Accuracy = 0.8259546630153659 Testing Accuracy = 0.8296577946768061
For Number of Estimator = 15 , Criterion = entropy , Maximum Depth = 30 :-
Training Accuracy = 1.0 Validation Accuracy = 0.8296059637912673 Testing Accuracy = 0.8293536121673004
For Number of Estimator = 20 , Criterion = entropy , Maximum Depth = 30 :-
Training Accuracy = 1.0 Validation Accuracy = 0.8273239008063289 Testing Accuracy = 0.831787072243346
```

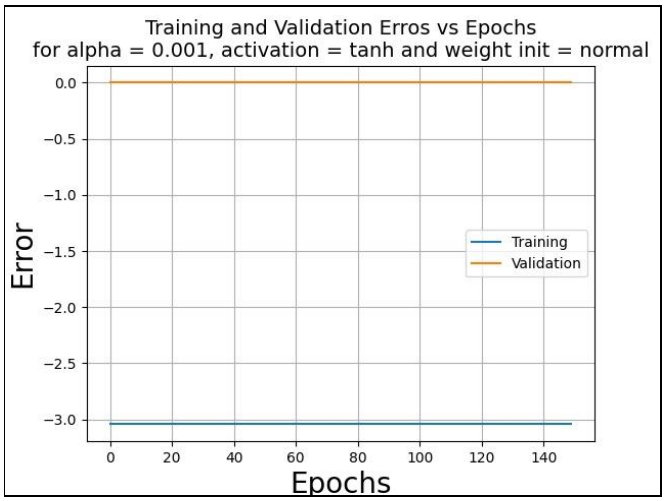
Q2

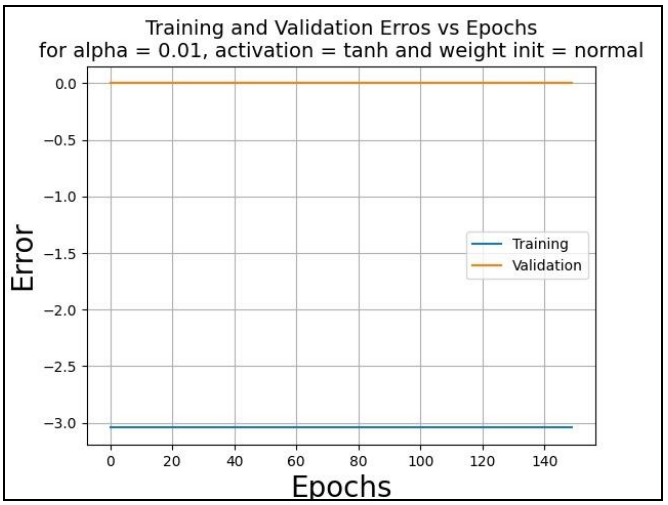
Explanation:

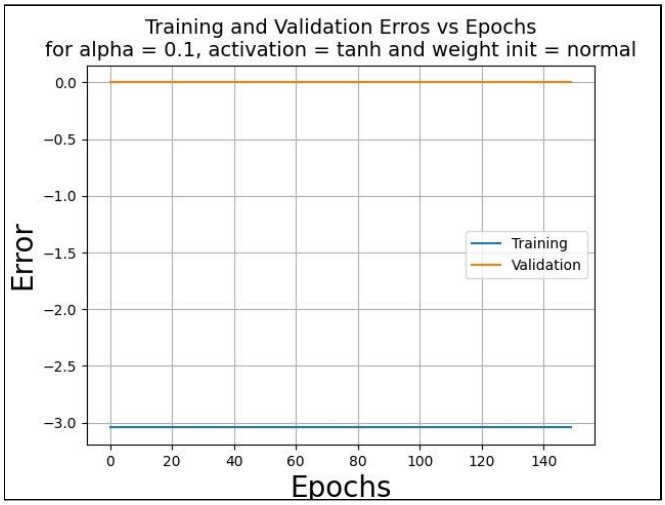
4.

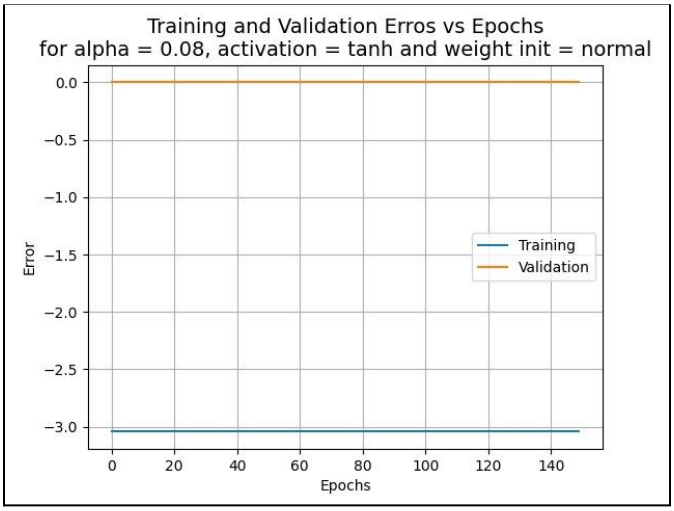
Sklearn’s sigmoid and tanh performs much better than mine implementation, because sklearn has been maintained and optimised over several years.

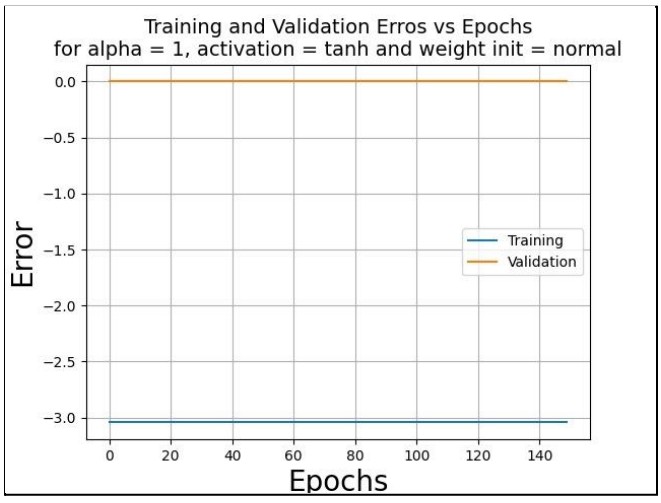
Plots:











Outputs:

1.

```
Testing Accuracy :  0.08571428571428572
```

4.

```
RELU : Training accuracy: 0.1273469387755102 Testing accuracy: 0.11142857142857143
Linear : Training accuracy: 0.10040816326530612 Testing accuracy: 0.08571428571428572
Sigmoid : Training accuracy: 0.5253061224489796 Testing accuracy: 0.5
tanh : Training accuracy: 0.47714285714285715 Testing accuracy: 0.46
```

5.

```
Testing Accuracy for learning rate = 0.001 and activation function used tanh :  0.08571428571428572
Testing Accuracy for learning rate = 0.01 and activation function used tanh :  0.08571428571428572
Testing Accuracy for learning rate = 0.1 and activation function used tanh :  0.08571428571428572
Testing Accuracy for learning rate = 1 and activation function used tanh :  0.08571428571428572
```

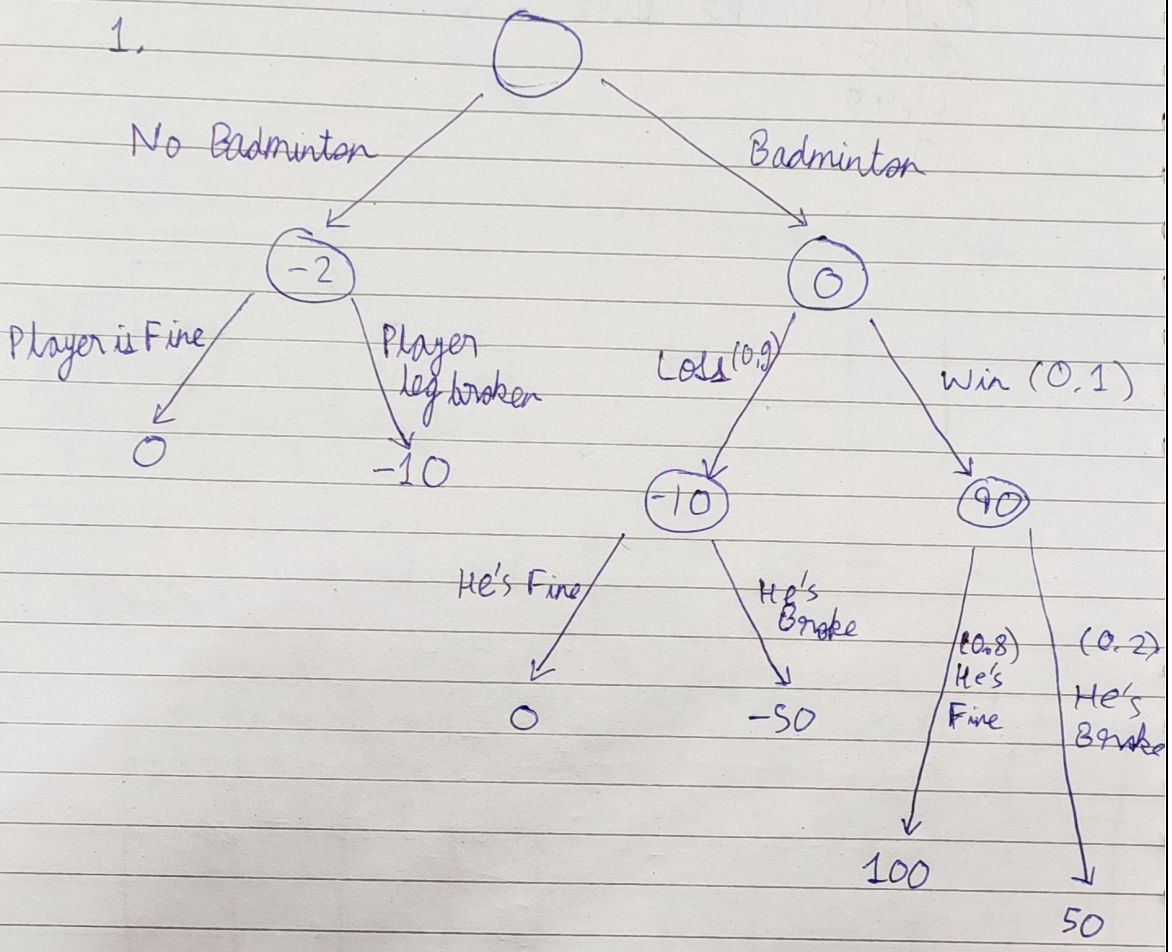
Theory:

Q1

Date : __/__/____

Theorey - 1

1.



2. From the decision tree, we can see that best action is to play badminton & provided he's fine and he also wins. In this case expected utility is 100
perfect

5. Yes, we can use the decision tree for finding whether ~~he~~ they will win the tournament depend on state of leg [Fine/Break]