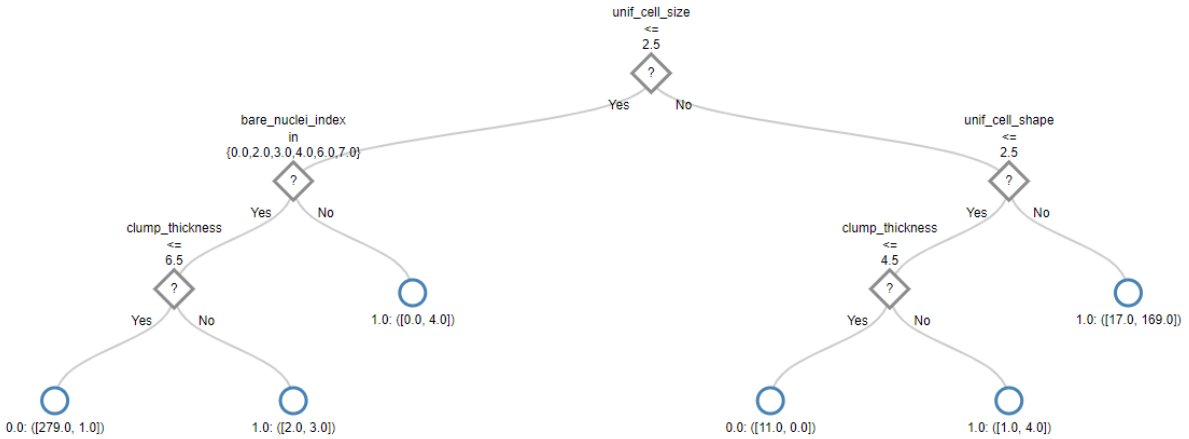
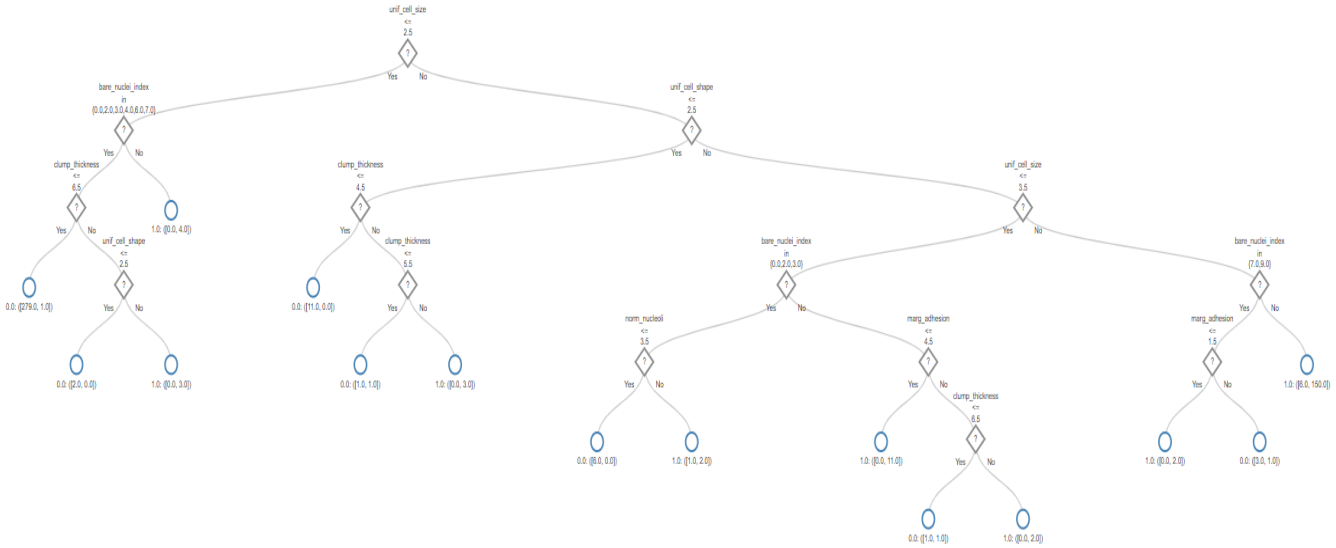
Kangdong Yuan

- (a) Show the decision tree visualization for the two values of max_depth (10 points)

The tree for max depth = 3



The tree for max depth = 6



- (b) Discuss the key difference between the two trees. (10 points)

The key difference between two trees is the depth of the tree. The depth of a decision tree is the length of the longest path from a root to a leaf.

- (c) Discuss which tree you believe is a better model for breast cancer diagnosis, and explain the rationale of your choice. (10 points)

I think second model (max_depth =6) is better for breast cancer diagnosis. Because I print the evaluator scores of two models, the second one has better score.

```
f1 = evaluator.evaluate(predictions_3)
print("f1 depth 3 score:", f1)
f1 = evaluator.evaluate(predictions_6)
print("f1 depth 6 score:", f1)
```

```
f1 depth 3 score: 0.9481309954934459
f1 depth 6 score: 0.9519230769230769
```

The depth of decision should fit to the number of features of prediction. If depth is too small it will cause underfitting. If depth is too large, it will cause overfitting. So, the max_depth = 3 cause underfitting, and max_depth = 6 is a suitable parameter.