Decision Trees

Procedure:

The continuous attributes of the input data were first converted into discrete by defining different ranges with the following length -

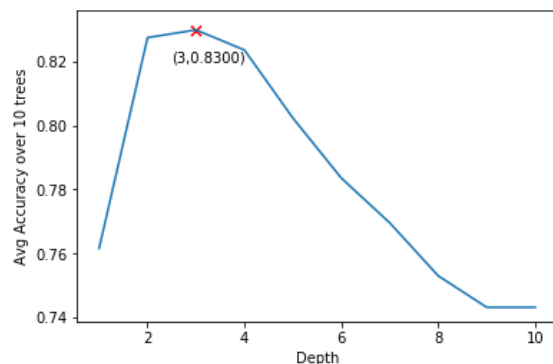| Age	10
Capital-gain	10000
Capital-loss	500
Hours-per-week	5
fnlwgt	100000

The length of the ranges solely on the basis of minimum and maximum of these values and human intuition.

Then on the preprocessed data, the ID3 algorithm was applied, as it is.

For each depth (1-10), 10 trees were formed using different random 80:20 splits in the available data. The average accuracy across these 10 trees was reported as the accuracy for that particular depth.

The depth for which the best accuracy was achieved (in our case, 3) was used to form the final tree. The available data was divided into 60:20:20, respectively for training, validation and testing. A new tree was trained for this depth, which was then post-pruned by Reduced Error Pruning method, using the validation set. A gain in accuracy on the test set was observed using the post pruned tree in comparison to the unpruned tree.

The plot of Depth vs Accuracy is attached below.

The final pruned decision tree:

{'Relationship': {'Husband': {'Education': {'10th': '<=50K',
                                            '11th': '<=50K',
                                            '12th': '<=50K',
                                            '1st-4th': '<=50K',
                                            '5th-6th': '<=50K',
                                            '7th-8th': '<=50K',
                                            '9th': '<=50K',
                                            'Assoc-acdm': {'Occupation': {'Adm-clerical': '<=50K',
                                                                          'Craft-repair': '<=50K',
                                                                          'Exec-managerial': '>50K',
                                                                          'Farming-fishing': '<=50K',
                                                                          'Handlers-cleaners': '<=50K',
                                                                          'Machine-op-inspct': '<=50K',
                                                                          'Other-service': '<=50K',
                                                                          'Prof-specialty': '<=50K',
                                                                          'Protective-serv': '>50K',
                                                                          'Sales': '>50K',
                                                                          'Tech-support': '>50K',
                                                                          'Transport-moving': '<=50K'}},
                                            'Assoc-voc': '<=50K',
                                            'Bachelors': {'hours_per_week': {0: '<=50K',
                                                                             5: '<=50K',
                                                                             10: '<=50K',
                                                                             15: '<=50K',
                                                                             20: '<=50K',
                                                                             25: '<=50K',
                                                                             30: '>50K',
                                                                             35: '>50K',
                                                                             40: '>50K',
                                                                             45: '>50K',
                                                                             50: '>50K',
                                                                             55: '>50K',
                                                                             60: '>50K',
                                                                             65: '>50K',
                                                                             70: '>50K',
                                                                             75: '>50K',
                                                                             80: '>50K',
                                                                             85: '<=50K',
                                                                             95: '>50K'}},
                                            'Doctorate': '>50K',
                                            'HS-grad': '<=50K',
                                            'Masters': {'hours_per_week': {0: '<=50K',

```
                                                    5: '>50K',
                                                    10: '<=50K',
                                                    15: '<=50K',
                                                    20: '<=50K',
                                                    25: '<=50K',
                                                    30: '>50K',
                                                    35: '>50K',
                                                    40: '>50K',
                                                    45: '>50K',
                                                    50: '>50K',
                                                    55: '>50K',
                                                    60: '>50K',
                                                    65: '>50K',
                                                    70: '<=50K',
                                                    75: '>50K',
                                                    85: '>50K',
                                                    95: '<=50K'}},
                        'Prof-school': '>50K',
                        'Some-college': {'Occupation': {'Adm-clerical': '<=50K',
                                                        'Armed-Forces': '>50K',
                                                        'Craft-repair': '<=50K',
                                                        'Exec-managerial': '>50K',
                                                        'Farming-fishing': '<=50K',
                                                        'Handlers-cleaners': '<=50K',
                                                        'Machine-op-inspct': '<=50K',
                                                        'Other-service': '<=50K',
                                                        'Prof-specialty': '<=50K',
                                                        'Protective-serv': '<=50K',
                                                        'Sales': '<=50K',
                                                        'Tech-support': '<=50K',
                                                        'Transport-moving': '<=50K'}}}},
    'Not-in-family': '<=50K',
    'Other-relative': '<=50K',
    'Own-child': '<=50K',
    'Unmarried': '<=50K',
    'Wife': {'Education': {'10th': '<=50K',
                          '11th': '<=50K',
                          '12th': '<=50K',
                          '5th-6th': '<=50K',
                          '7th-8th': '<=50K',
                          '9th': '<=50K',
                          'Assoc-acdm': {'hours_per_week': {0: '>50K',
                                                            10: '>50K',
                                                            15: '<=50K',
```

```
                     20: '<=50K',
                     25: '>50K',
                     30: '>50K',
                     35: '<=50K',
                     40: '<=50K',
                     45: '>50K',
                     50: '>50K',
                     55: '<=50K',
                     70: '<=50K',
                     95: '<=50K'}},
'Assoc-voc': {'Age': {25: '<=50K',
              35: '>50K',
              45: '>50K',
              55: '<=50K',
              65: '>50K',
              75: '<=50K'}},
'Bachelors': {'native_country': {'Canada': '<=50K',
                  'China': '<=50K',
                  'Cuba': '>50K',
                  'England': '>50K',
                  'Germany': '>50K',
                  'Haiti': '>50K',
                  'Hong': '<=50K',
                  'Japan': '>50K',
                  'Mexico': '<=50K',
                  'Philippines': '>50K',
                  'Puerto-Rico': '<=50K',
                  'South': '<=50K',
                  'Taiwan': '<=50K',
                  'Thailand': '<=50K',
                  'United-States': '>50K'}},
'Doctorate': '>50K',
'HS-grad': '<=50K',
'Masters': '>50K',
'Prof-school': '>50K',
'Some-college': '<=50K'}}}}
```