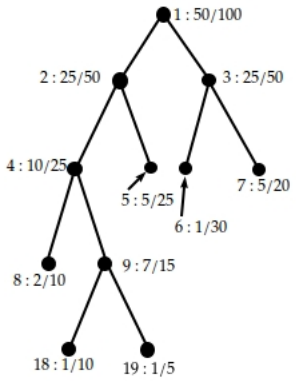
B365 Homework 6

Zac Monroe

November 2018

1. (a) This is my fully-deepened tree.



(b) $R(T) = \frac{5}{25}\frac{25}{100} + \frac{1}{30}\frac{30}{100} + \frac{5}{20}\frac{20}{100} + \frac{2}{10}\frac{10}{100} + \frac{1}{10}\frac{10}{100} + \frac{1}{5}\frac{5}{100} = \frac{15}{100} = 0.15$

(c) I do not believe that this value of $R(T) = 0.15$ is an accurate representation of how the tree would perform on new data. The tree splits many times for a rather small data set, and many terminal nodes have very few data points to be classified, thus I believe that the tree is over-fitting to the training data.

(d) First, compute R^*_α ($\alpha = 0.04$) for terminal nodes:

(5) $R^*_\alpha(5) = \frac{5}{25}\frac{25}{100} = 0.05$

(6) $R^*_\alpha(6) = \frac{1}{30}\frac{30}{100} = 0.01$

(7) $R^*_\alpha(7) = \frac{5}{20}\frac{20}{100} = 0.05$

(8) $R^*_\alpha(8) = \frac{2}{10}\frac{10}{100} = 0.02$

(18) $R^*_\alpha(18) = \frac{1}{10}\frac{10}{100} = 0.01$

(19) $R^*_\alpha(19) = \frac{1}{5}\frac{5}{100} = 0.01$

Next, for non-terminal nodes:

(9) $R^*_\alpha(9) = min(\frac{7}{15}\frac{15}{100},\ \alpha + R^*_\alpha(18) + R^*_\alpha(19)) = min(0.07,\ 0.04 + 0.01 + 0.01) = 0.06$

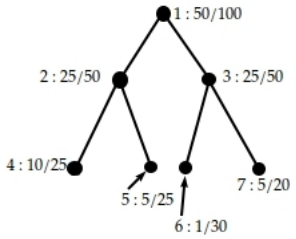
(3) $R^*_\alpha(3) = min(\frac{25}{50}\frac{50}{100},\ \alpha + R^*_\alpha(6) + R^*_\alpha(7)) = min(0.25,\ 0.04 + 0.01 + 0.05) = 0.11$

(4) $R^*_\alpha(4) = min(\frac{10}{25}\frac{25}{100},\ \alpha + R^*_\alpha(8) + R^*_\alpha(9)) = min(0.10,\ 0.04 + 0.02 + 0.06) = 0.10 \rightarrow$ prune

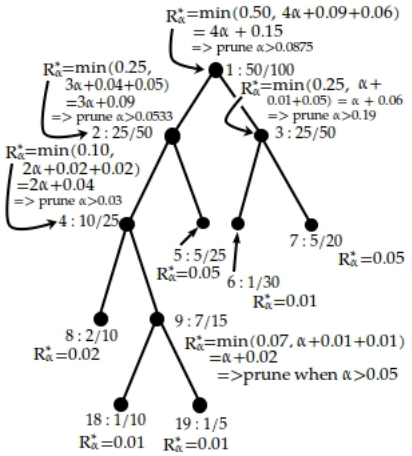
(2) $R^*_\alpha(2) = min(\frac{25}{50}\frac{50}{100},\ \alpha + R^*_\alpha(4) + R^*_\alpha(5)) = min(0.25,\ 0.04 + 0.10 + 0.05) = 0.19$

(1) $R^*_\alpha(1) = min(\frac{50}{100}\frac{50}{100},\ \alpha + R^*_\alpha(2) + R^*_\alpha(3)) = min(0.50,\ 0.04 + 0.06 + 0.11) = 0.21$

Thus T_α for $\alpha = 0.04$ is this:

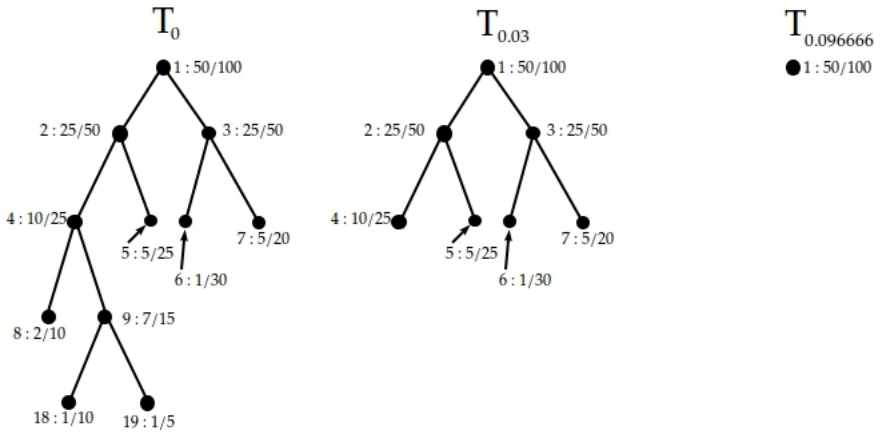


(e) Here is the R^*_α value for each node when $\alpha = 0$.



We can see that node 4 would be the first to be pruned because it has the next lowest α-value that would make it be pruned (any $\alpha > 0.03$).

(f) Here is the family of nested subtrees for each value of α.



2. (a) Source prob2.r into R to use my optimal risk function. optimal_risk takes in one integer in range [1,19] and outputs the optimal risk for that node in the tree.

(b) I would construct this tree by first having the initial tree T_0 and pruning any non-terminal node i for which $p(i)p(error|i) < 0.04 + \sum_{i \to j} R^*_\alpha(j)$ (where $i \to j$ denotes all children nodes j of node i). This computation is done by the recursive R^*_α function.

(c) A problem that has an end goal of classification of a couple types may want to use a decision tree. A problem that doesn't want to over-fit their tree to the training data would use $T_{\alpha=0.04}$ instead of just

the root node T_0 or because there is still a desire for decent proper classification (just using the root node would give a poor error rate) while looking to minimize complexity (using T_0 is the most complex decision tree that can be built out of the training data; $T_{\alpha=0.04}$ is less complex while still having some depth for decisions to be made about new test data).

3. (a) Having a rel error value of 0 indicates that this tree performed the best on the training data relative to all the other trees in $\{T_\alpha\}$. It made no mistakes on the training data.

   (b) I think this tree will perform very poorly on novel data because it is extremely complex and has likely over-fit itself onto the data. New data points may be placed in the wrong class, even if training points were all placed in the correct one.

   (c) That tree will likely classify new data properly with the same error rate that it had on the training data, because the most likely class will probably not change in its representation in the corpus as a whole.

   (d) The best choice of complexity parameter α seems to be $\alpha = 0.00064725$ because the associated decision trees made with the data using this complexity parameter have the lowest cross-validation error (i.e. they perform best on novel data), and that cross-validation error thus seems to be a minimum value for each α-value.

4. (a) Look at prob4.r to see my code filling in the array z.

   (b) Source prob4.r into R and test out understands(), passing in a vector of length 7 where each element in the vector is either 0 or 1.