Some problems with Decision Trees

(Konstantin Burlachenko, 31JULY2018, burlachenkok@gmail.com)

Decision Tree Problems

The deep reason for bagging and various ensemble technics that this technic solve one of the problem with the trees – the problem that decision trees have high variance due to four reasons:

1. The possible error in upper part of decision tree is propagate down to the terminal nodes. The error are not accumulated (or averaged), but in fact error is cascaded if look into how decision are making from the top part of the tree to the bottom level. And due to that all structure of decision tree is building with random samples they are not infinite, and even at each level of the tree "number of samples" is reduced.

2. The decision tree is building with samples from population. Small change in this train set will lead to big change:
   a. Of the tree structure
   b. Tree construction is very unstable
   c. High variance due to sample fluctuation

3. Even minor change split in the root will dramatically change tree structure

4. Tree's prediction involve multiplication of the indicator functions if look into tree structure. And "error" that have been made are multiplied. It's rather bad compare to situation when "errors" are summed. Because in sum are some errors can cancel each other.

"The reasons of all of this is that in non-convex optimization beside local minimums there is another problems with unstable solution due random sample" – Jerome H.Friedman.

Can decision Tree Problems can be fixed by Decision Tree model?

No. There are currently couple state-of-art models for decision trees:

CART and C4.5/C5.0 they have minor difference in:

1. How classification is performed
2. Is implementation opensource or not. C5.0 – is commercial version of C4.5

What todo with Decision Tree problems

(1) Live with the problem
(2) Use other methods for make predictions
(3) Fixup trees. Recent research is:
   a. Bagging 1996 by Leo Brieman
   b. Boosting 1996 by Leo Brieman and Jerome H.Friedman
   c. MARS (Multiple Adaptive Regression Splines) 1989.