2.3 Methods of construction of decision trees.

Existing methods (there are dozens of methods) can be divided into two basic groups. The first group is methods of building of a strict-optimum tree (by the given criteria of quality of a tree) and the second group is methods of construction of an approximately optimum tree.

The problem of searching the optimum variant of a tree can be related to a discrete programming problem or a choice from finite (but very large) numbers of variants. It follows from the fact that for finite training sample the number of variants of branching (see below) for each characteristic is finite.

Three basic kinds of methods in discrete programming are considered: exhausting search, a method of dynamic programming and a branch and bounds method. However, these methods, in the application to decision trees, as a rule, are very laborious, especially for the large number of observations and characteristics. Therefore, we will consider approximate methods: method of consecutive branching, a method of pruning and a recursive method.

Let us consider all basic operations with decision trees. Methods of tree constructing will represent the certain sequence of these operations.

2.3.1 Operation of branching (division).

This operation is the basic operation for tree constructions. We will consider a node of a tree and some characteristic X_j. Let the range of definition of this characteristic be divided on L_j subsets (ways of a choice of such subsets we will consider below). In case of the quantitative characteristic, these subsets represent a set subintervals of splitting, in case of the qualitative characteristic a subsets of values and in case of the ordered characteristic the subsets including the neighbouring values.

Let us associate with each of these subsets a branch of the tree leaving the given (parent) node and going into a new node which is called descendant. Thus, the node "has branched" ("has divided") on L_j new nodes (figure 5).

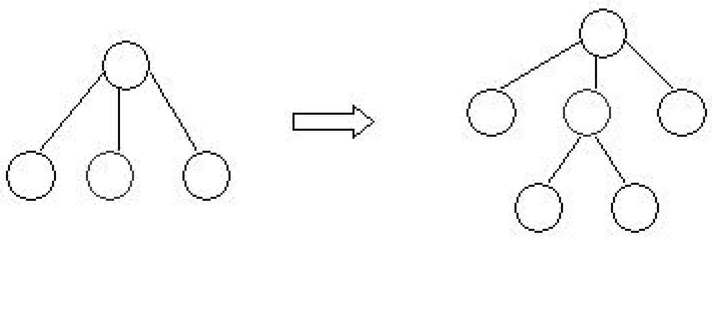

Fig. 5

Notice that for binary trees L_j is always equal to two. If L_j is always equal to three such trees are called *ternary*. If L_j is always equal to four we receive quadratic-trees.

How to obtain the splitting of a range of definition? We take a set of the observations corresponding to the given node and consider values of characteristic X_j for these observations.

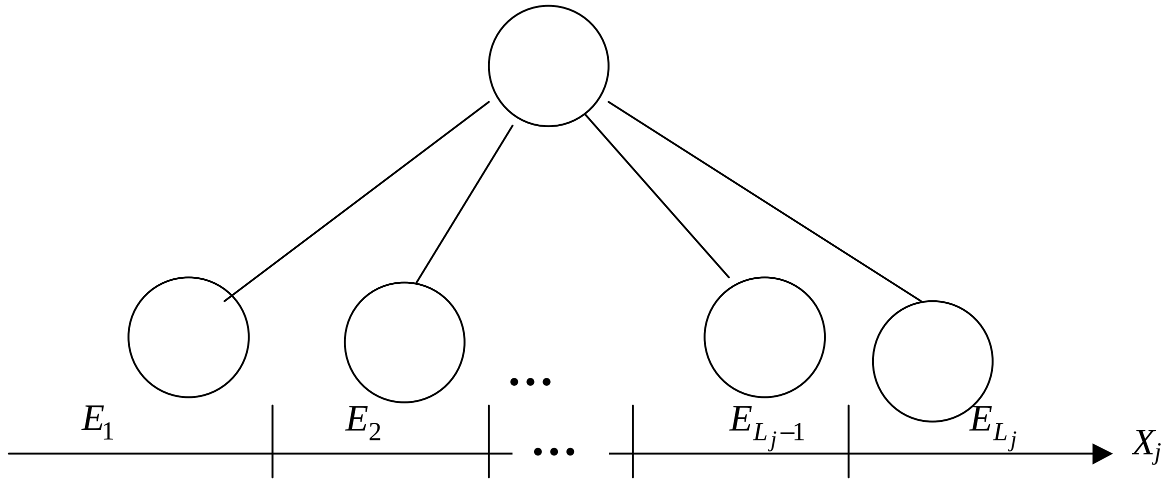

Fig. 6

Consider a quantitative characteristic. In this case, boundaries are in the middle of intervals between the neighbor values and splitting carried out on these boundaries (figure 6).

For example on figure 7 (values of the characteristic for observations are designated through ⊗), in case of a binary tree, it is possible to consider the following variants of splitting: X_j < 0.5 or X_j >=0.5, X_j < 1.5 or X_j >=1.5. If the characteristic is qualitative, then the variants of splitting are values of the characteristic, for example if X_j means a country, the following splitting can be received: $X_j \in$ {Canada, Mexico, USA} or $X_j \in$ {Argentina, Brazil}.

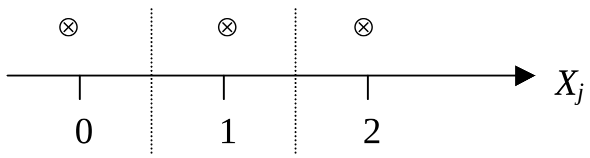

Fig. 7

In case of the large number of values, the number of possible variants of splitting becomes too big, therefore for acceleration of process of tree construction one considers not all variants, but only some of them (for example, such as X_j =”Canada» or $X_j \neq$ “Canada»).

In case of the ordered characteristic, variants of splitting consist of the ordered values, for example if X_j is a military rank, division can be such: $X_j \in$ [the soldier – the first sergeant] or $X_j \in$ [the lieutenant - the major].

For qualitative or ordered characteristics, it can happen (when sample size of observations is small) that the set of values of characteristics for the observations appropriate to the node, are only a part of all range of definition of this characteristic. In this case, it is necessary to attribute the rest values to a new branch. Now at forecasting the object of the control sample, which had such value, we can define to which branches it belongs. For example, it is possible to attribute the given values to the branch, corresponding with the greatest number of observations.

2.3.2 Operation of the definition of degree of promise for branching node (rule of a stopping).

Let us consider a dangling node of a tree, i.e. the node which is not branched, but it is not clear, whether this node will be a leaf or whether we need further branching. We will consider the subset of observations appropriate to the given node. We will divide the nodes into two cases. First, if these observations are homogeneous, i.e. basically belong to the same class (a pattern recognition problem, RP), or if the variance of *Y* for them is small enough (a regression analysis problem, RA). The variant when the values of characteristic are equal for all observations corresponds also to this case. Second, if the number of observations is not enough.

The node, unpromising for the further branching, is called leaf.

For definition of degree of promise, it is possible to set the following parameters: an allowable error for node (PR problem), an allowable variance (RA problem) and a threshold on the quantity of observations.

2.3.3 Operation "to attribute a solution to a leaf ".

Let us consider a leaf of a tree. A subset of observations *Data* corresponds to this leaf. During the solution of a pattern recognition problem, the class with maximal quantity of observations from *Data* is assigned to a leaf, in comparison with other classes. At the solution of a regression analysis problem, the solution attributed to a leaf, is equal to the average value of dependent characteristic *Y* for observation from *Data*.

2.3.4. Operation of "growth" of node.

This operation represents a sequence of operations of branching for each of the new nodes of a tree. As a result of this operation, the given node is replaced with some sub tree (i.e. a part of a full tree which also looks like a tree (figure 8)).

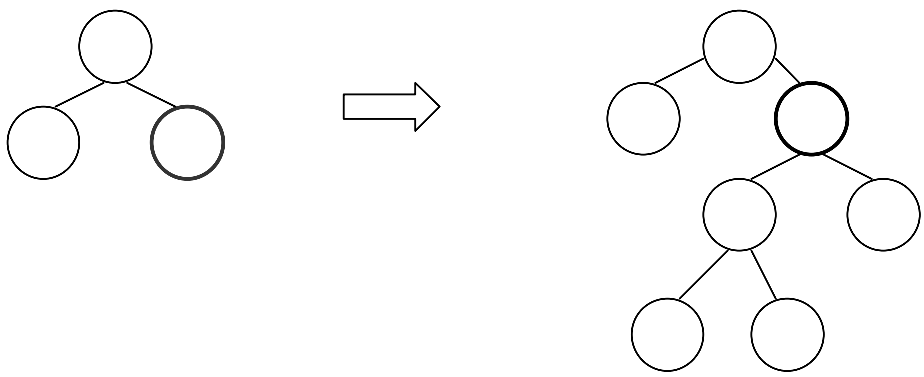

Fig. 8

The complexity of the sub tree is limited to a parameter. One of the ways of growth will be described below more detailed.

2.3.5. Operation of pruning.

This operation is the opposite of operations of growth, i.e. for the given node appropriate sub tree for which this node is a root completely cuts (figure 9). The node is then called a leaf.

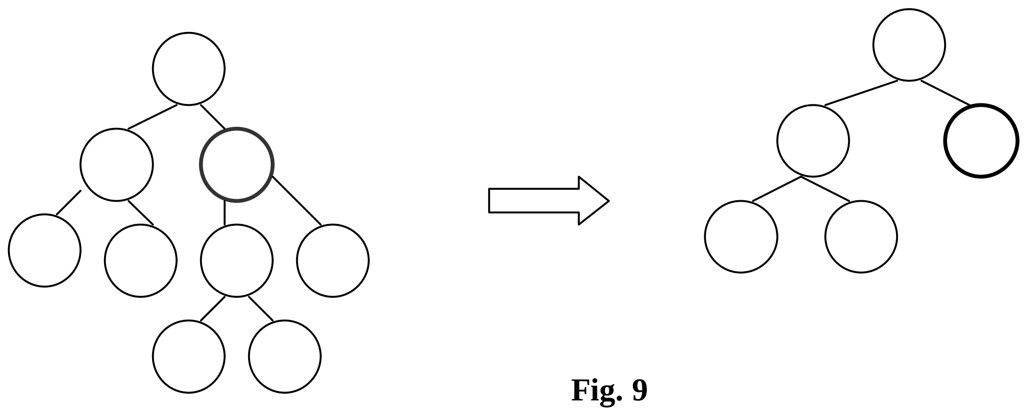

Fig. 9

2.3.6. Operation of "aggregation" of nodes or ("join").

Let the node be divided on *L* new nodes. We will take any pair of these nodes and we will unite them in one node and connect it with the parental node (figure 10). Thus subsets of the values to the appropriate aggregation nodes are united.

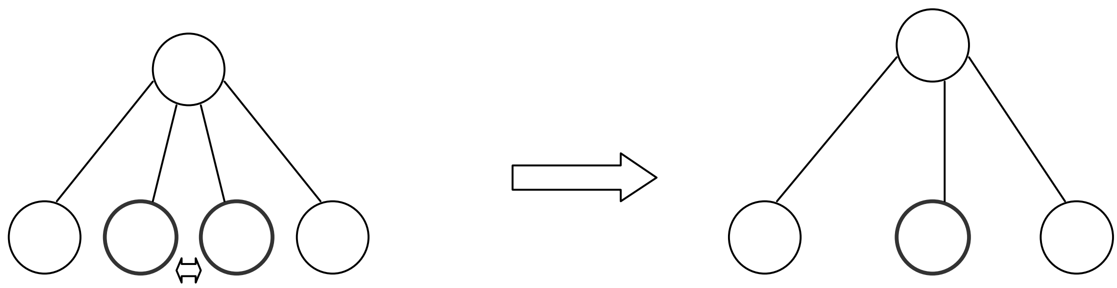

Fig.10

The aggregation of nodes, which correspond to the neighbor subintervals or subsets of values of quantitative and ordered characteristics is allowed.