

# Numeric Feature Splits Algorithm

// Run this subroutine for each numeric feature at each node of DT induction

DetermineCandidateNumericSplits(set of training instances  $D$ , feature  $X_i$ )

$C = \{\}$  // initialize set of candidate splits for feature  $X_i$

let  $v_j$  denote the value of  $X_i$  for the  $j^{th}$  data point

sort the dataset using  $v_j$  as the key for each data point

for each pair of adjacent  $v_j, v_{j+1}$  in the sorted order

if the corresponding class labels are different

add candidate split  $X_i \leq (v_j + v_{j+1})/2$  to  $C$

return  $C$