C 4.5 Source code Pt.3

ISL / 강한솔

Index

- ✓ Tree structure
- ✓ Build.h
- ✓ Tree.h
- ✓ St-thresh.h

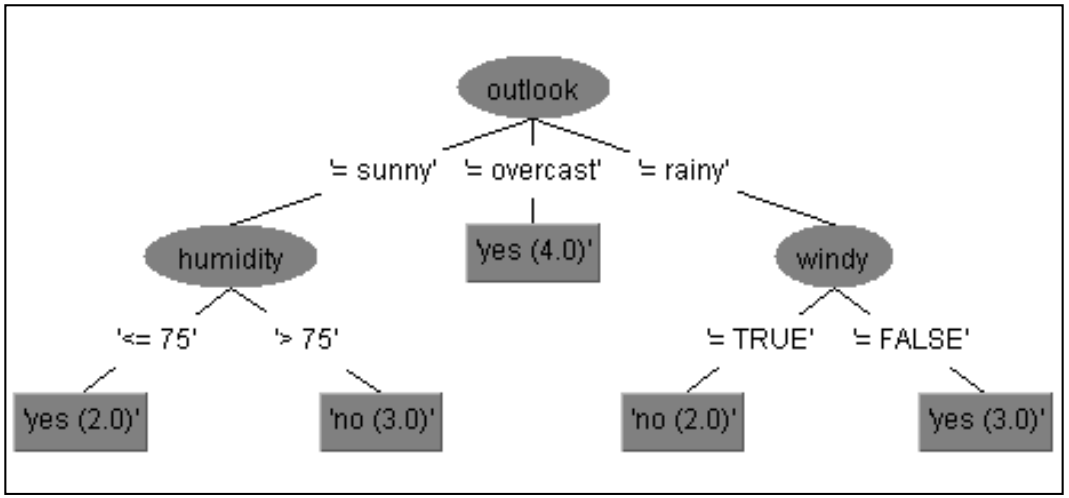
Tree structure

*Concpets : Node, Branch, Leaf, Subtree, Attribute, Attribute Value, Class

Play, Don't Play.

outlook: sunny, overcast, rain.
temperature: continuous.
humidity: continuous.
windy: true, false.

<File Format>



<Decision Tree>

short	NodeType;	// 0=leaf 1=branch 2=cut 3=subset
ClassNo	Leaf;	//현재 노드에서 평가하는 클래스 인덱스
ItemCount	Items,	//현재 노드에서의 항목 수
	*ClassDist,	//현재 노드에서 평가하는 클래스의 수
	Errors;	//현재 노드에서의 에러 수
Attribute	Tested;	//현재 테스트 하고 있는 속성의 인덱스
short	Forks;	//현재 노드에서 branch 수
float	Cut,	//threshold 값
	Lower,	//낮은 threshold
	Upper;	//높은 threshold
Set	*Subset;	//명목형 속성의 subset
Tree	*Branch;	//Branch[x] subtree집합

<Tree Structure>

ISL Image System Laboratory

Build.h

- void InitialiseTreeData()

  트리 작성과 관련된 메모리를 할당한다.

- void InitialiseWeight()

  가중치를 1로 초기화한다.

- Tree FormTree(ItemNo Fp, ItemNo Lp)

  Decision tree를 만들어준다.

- ItemNo Group (DiscrValue V, ItemNo Fp, IemNo Lp, Tree TestNode)

  서로 연관성 있는 항목끼리 그룹화 시켜준다.

ISL Image System Laboratory

Build.h

- ItemCount CountItems(ItemNo Fp, ItemNo Lp)

  가중치를 둬서 항목들의 개수를 구한다.

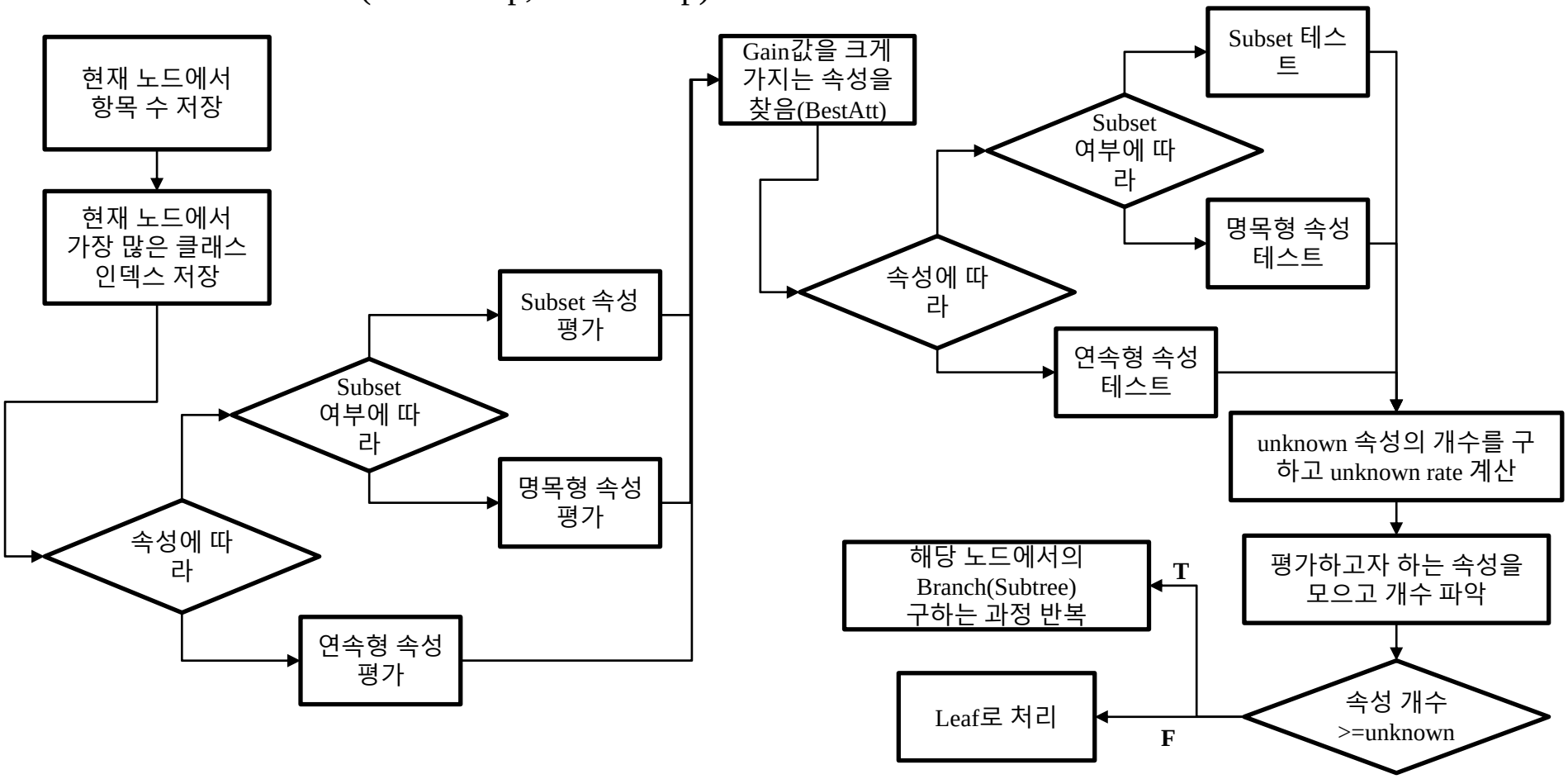
- void Swap (ItemNo a, ItemNo b)

  a 인덱스를 가지는 항목과 b 인덱스를 가지는 항목을 바꾼다.

ISL Image System Laboratory

Build.h

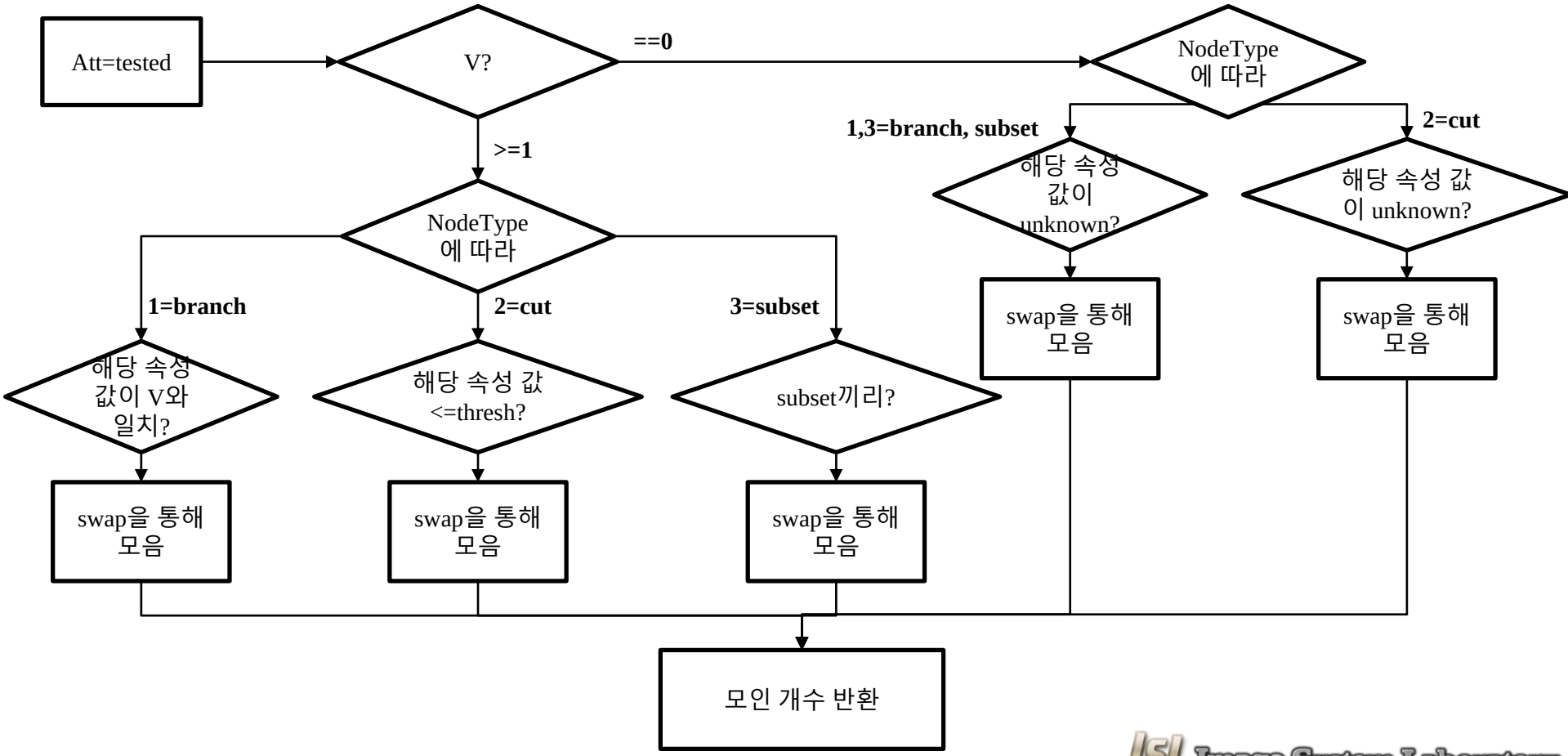
✓ Tree FormTree(ItemNo Fp, ItemNo Lp)



ISL Image System Laboratory

Build.h

- ItemNo Group (DiscrValue V, ItemNo Fp, IemNo Lp, Tree TestNode)



ISL Image System Laboratory

Tree.h

- void PrintTree(Tree T)

  Decision Tree를 출력한다.

- void Show(Tree T, short Sh)

  PrtintTree에서 호출되는 함수로, 실질적인 출력하는 역할(leaf)을 한다.

- void ShowBranch(short Sh, Tree T, DiscrValue v)

  Show에서 호출되는 함수로 Branch를 출력하는 함수다

- short MaxLine(Tree St)

  leaf가 없는 하위 subtree의 한 줄의 최대 크기를 알려주는 함수다.

ISL Image System Laboratory

Tree.h

- void Indent(short Sh, char *Mark)

  들여쓰기해주는 함수로 branch를 가시적으로 나타내기 위해 사용된다.

- void SaveTree(Tree T, String Extension)

  OutTree를 호출해 Decision Tree를 저장하는 함수다.

- void OutTree(Tree T)

  Tree를 char 형태의 데이터로 저장하는 함수다.(바이너리 파일과 흡사한 형태)

- Tree GetTree(String Extension)

  SaveTree와 반대로 InTree를 호출해 Decision Tree를 읽어오는 함수다.

ISL Image System Laboratory

Tree.h

- Tree InTree()

  OutTree와 반대로 char형태로 저장된 Tree를 읽어오는 형태의 함수다.

- void StreamOut(String s, int n)

  OutTree에서 사용되는 함수로 char형태로 저장한다.

- void StreamIn(String s, int n)

  InTree에서 사용되는 함수로 char형태로 저장된 형태를 다시 tree형태로 복원한다.

- void ReleaseTree(Tree Node)

  해당 노드에 의해 잡힌 메모리를 해제 하는 함수다.

ISL Image System Laboratory

Tree.h

- Tree Leaf(ItemCount *ClassFreq, ClassNo NodeClass, ItemCount Cases, ItemCount Errors)

  주어진 노드의 leaf를 구하는 함수다.

- void Sprout(Tree Node, DiscrValue Branches) :

  해당 노드에 branch를 추가하는 함수다.

- int TreeSize(Tree Node) :

  노드의 개수를 세는 함수이다.

- Tree CopyTree(Tree T)

  Tree를 복사하는 함수다.

ISL Image System Laboratory

Tree.h

- void SaveDiscreteNames()

  명목형 속성들을 char형태로 저장한다.

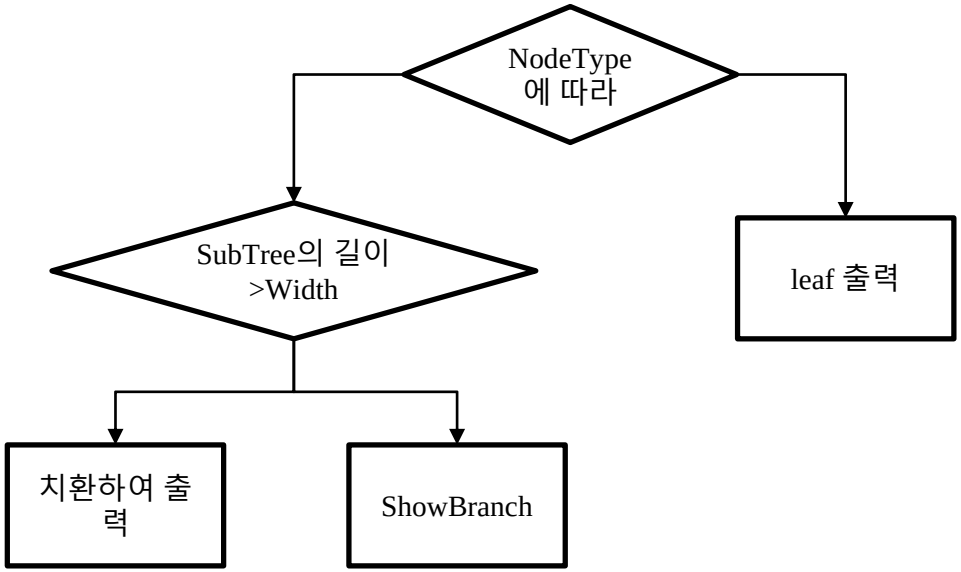
- void RecoverDiscreteNames()

  char형태로 저장된 명목형 속성들을 다시 복원한다.

ISL Image System Laboratory

Tree.h

- ✓ void Show(Tree T, short Sh)




```
Decision Tree:

outlook = overcast: Play (4.0)
outlook = sunny:
|   humidity <= 75 : Play (2.0)
|   humidity > 75 : Don't Play (3.0)
outlook = rain:
|   windy = true: Don't Play (2.0)
|   windy = false: Play (3.0)
```

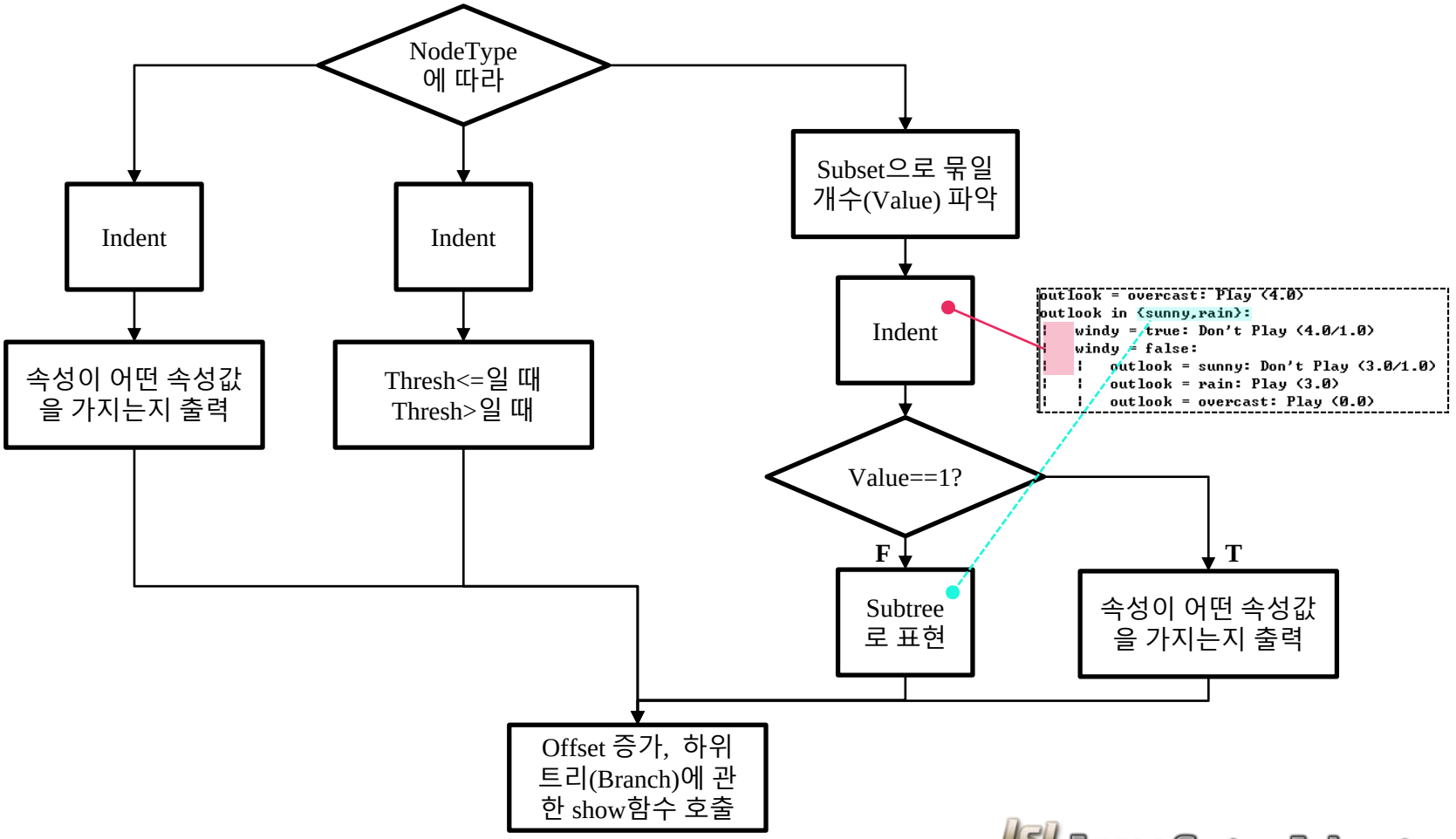
⇩

```
Decision Tree:

outlook = overcast: Play (4.0)
outlook = sunny:[S1]
outlook = rain:[S2]


Subtree [S1]

humidity <= 75 : Play (2.0)
humidity > 75 : Don't Play (3.0)


Subtree [S2]

windy = true: Don't Play (2.0)
windy = false: Play (3.0)
```


ISL Image System Laboratory

Tree.h

- ✓ void ShowBranch(short Sh, Tree T, DiscrValue v)



ISL Image System Laboratory

St-thresh.h

- void SoftenThresh(Tree T)

  연속형 속성에 대한 Thresholds값을 부드럽게 하는 함수다.

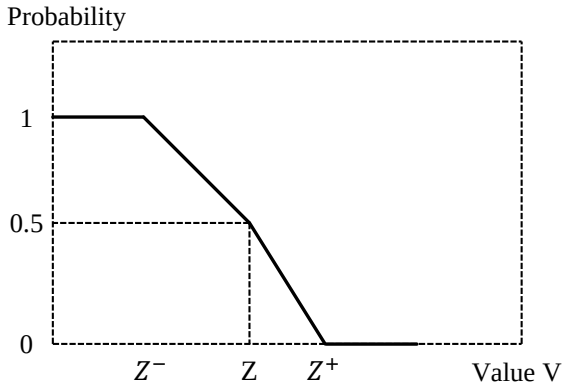
- void ScanTree(Tree T, ItemNo Fp, ItemNo Lp)

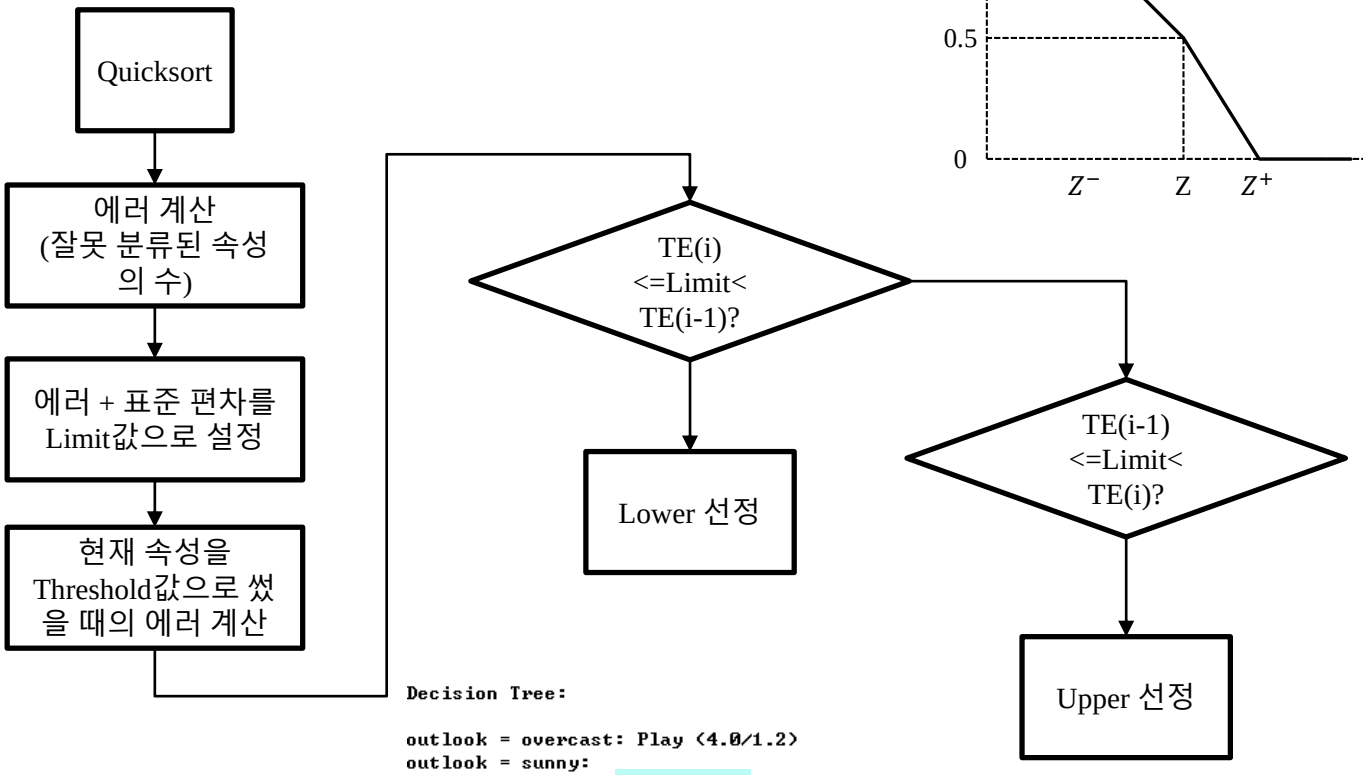
  연속형 속성에서 나누는 기준(upper, lower)을 계산하는 함수다.

ISL Image System Laboratory

St-thresh.h

✓ void ScanTree(Tree T, ItemNo Fp, ItemNo Lp)






```
Decision Tree:

outlook = overcast: Play (4.0/1.2)
outlook = sunny:
|   humidity <= 75 [70,80.155]: Play (2.0/1.0)
|   humidity > 75 [70,80.155]: Don't Play (3.0/1.1)
outlook = rain:
|   windy = true: Don't Play (2.0/1.0)
|   windy = false: Play (3.0/1.1)
```


ISL Image System Laboratory

Q & A