CLINICAL INTERPRETATION OF OMICS CLUSTERING RESULTS

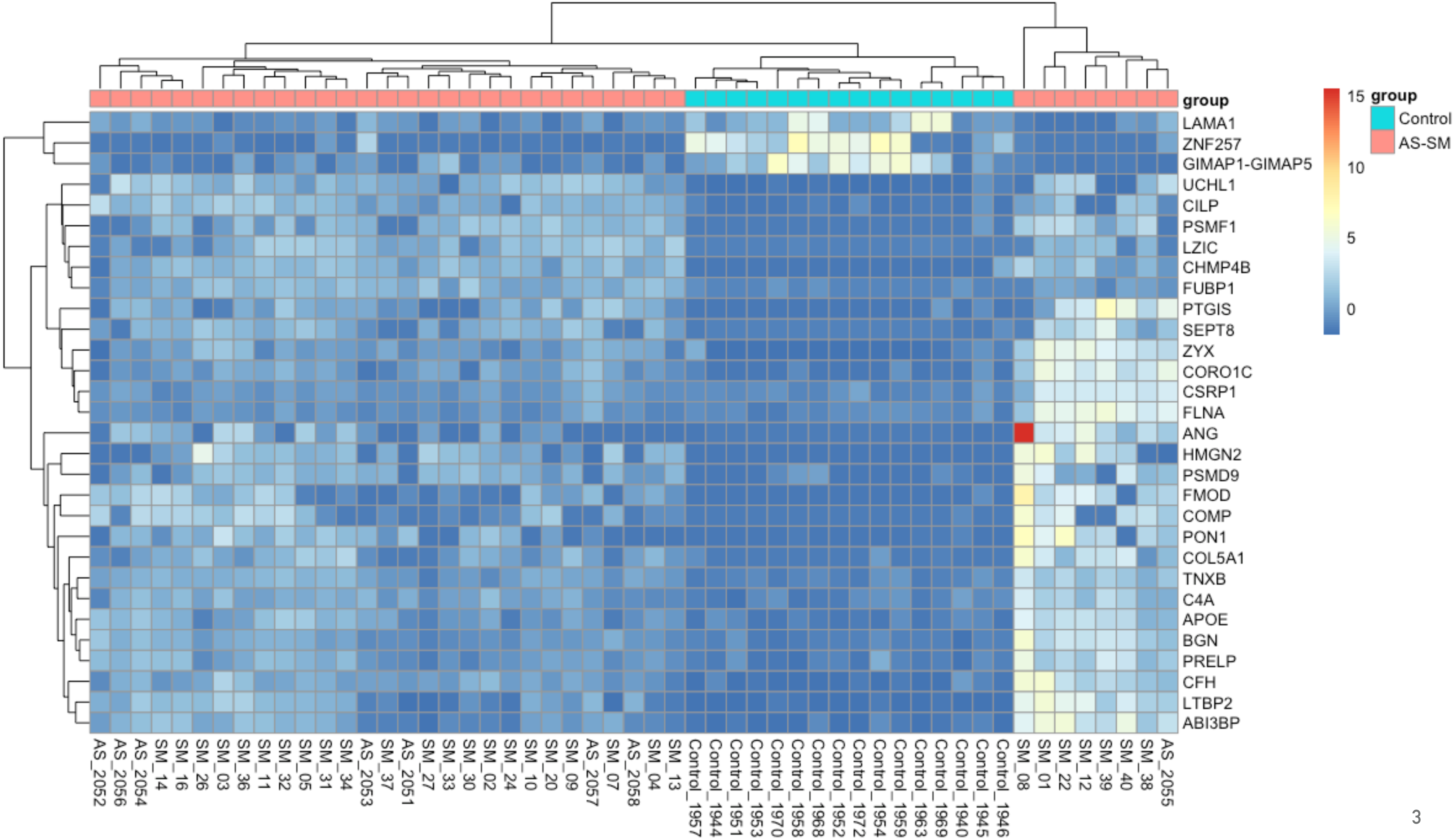
Trends in Bioinformatics WS 17/18

Supervisor: Milena Kraus

Ajay Kesar

AGENDA

- Task
- Data
- Feature Selection Models
  - Filter / Wrapper (DT) / Embedded
- Decision Tree Baseline
- Regularized Tree Model
- Future Work


group
15
10
5
0
group
Control
AS-SM
LAMA1
ZNF257
GIMAP1-GIMAP5
UCHL1
CILP
PSMF1
LZIC
CHMP4B
FUBP1
PTGIS
SEPT8
ZYX
CORO1C
CSRP1
FLNA
ANG
HMGN2
PSMD9
FMOD
COMP
PON1
COL5A1
TNXB
C4A
APOE
BGN
PRELP
CFH
LTBP2
ABI3BP
AS_2052
AS_2056
AS_2054
SM_14
SM_16
SM_26
SM_03
SM_36
SM_11
SM_32
SM_05
SM_31
SM_34
AS_2053
SM_37
AS_2051
SM_27
SM_33
SM_30
SM_02
SM_24
SM_10
SM_20
SM_09
AS_2057
SM_07
AS_2058
SM_04
SM_13
Control_1957
Control_1944
Control_1951
Control_1953
Control_1970
Control_1958
Control_1968
Control_1952
Control_1972
Control_1954
Control_1959
Control_1963
Control_1969
Control_1940
Control_1945
Control_1946
SM_08
SM_01
SM_22
SM_12
SM_39
SM_40
SM_38
AS_2055

A	B	C	D	E	F	G	H
	PatientID	Category	ParameterName	Event	Value	Unit	Unnamed: 6
0	SM 01	Demographics / Clinical Parameters	Gender	General	female		
1	SM 01	Demographics / Clinical Parameters	Study group	General	No cardiac medication		
2	SM 01	General	Time between VISIT 1 and VISIT 2	General	223		
3	SM 01	General	Time between VISIT 1 and surgery	General	3		
4	SM 01	General	Time between surgery and VISIT 2	General	220		
5	SM 01	Surgery Parameters	Age at Surgery	Surgery	69	years	
6	SM 01	Surgery Parameters	Aortic Valve Replacement	Surgery	yes		
7	SM 01	Surgery Parameters	MIC	Surgery	yes		
8	SM 01	Surgery Parameters	Additional Surgery	Surgery	no		
9	SM 01	Surgery Parameters	CABG Surgery	Surgery	no		
10	SM 01	Surgery Parameters	Other Surgery	Surgery	no		
11	SM 01	Surgery Parameters	Aortic Valve Size	Surgery	23	mm	
12	SM 01	Surgery Parameters	Biological Aortic Valve?	Surgery	yes		
13	SM 01	Surgery Parameters	Modell of Aortic Valve	Surgery	CE Perimount Magna Ease 23mm, 3300 TFX		
14	SM 01	Surgery Parameters	Aortic Clamping Time	Surgery	87	min	
15	SM 01	Surgery Parameters	Perfusion Time	Surgery	122	min	
16	SM 01	Surgery Parameters	Reperfusion Time	Surgery	18	min	
17	SM 01	Surgery Parameters	Cardiac Arrest Time	Surgery	87	min	
18	SM 01	Surgery Parameters	Biopsy	Surgery	yes		
19	SM 01	Surgery Parameters	Postoperative Pacemaker	Surgery	no		
20	SM 01	Catheter Measurement	Left ventricular systolic pressure	General	230	mm[Hg]	2015-03-18
21	SM 01	Catheter Measurement	Left ventricular end diastolic pressure	General	15	mm[Hg]	2015-03-18
22	SM 01	Catheter Measurement	Pressure in the ascending aorta	General	180	mm[Hg]	2015-03-18
23	SM 01	Demographics / Clinical Parameters	Height	VISIT 1	164	cm	
24	SM 01	Demographics / Clinical Parameters	Weight	VISIT 1	71	kg	
25	SM 01	Demographics / Clinical Parameters	NYHA stage	VISIT 1	2		
26	SM 01	Demographics / Clinical Parameters	Blood pressure systolic right arm	VISIT 1	164	mm[Hg]	
27	SM 01	Demographics / Clinical Parameters	Blood pressure diastolic right arm	VISIT 1	80	mm[Hg]	
28	SM 01	Demographics / Clinical Parameters	Blood pressure mean right arm	VISIT 1	122	mm[Hg]	
29	SM 01	Demographics / Clinical Parameters	Blood pressure systolic left arm	VISIT 1	164	mm[Hg]	
30	SM 01	Demographics / Clinical Parameters	Blood pressure diastolic left arm	VISIT 1	80	mm[Hg]	
31	SM 01	Demographics / Clinical Parameters	Blood pressure mean left arm	VISIT 1	122	mm[Hg]	
32	SM 01	Demographics / Clinical Parameters	Time of blood pressure	VISIT 1	5	pm	
33	SM 01	Demographics / Clinical Parameters	Heart rate	VISIT 1	76	per minute	
45	SM 01	Diagnoses	Diagnoses		AS III		
46	SM 01	Diagnoses	Diagnoses		Hypercholesterinemia, (fam.)		
47	SM 01	Diagnoses	Diagnoses		Gastritis (Autoimmune)		
48	SM 01	Diagnoses	Diagnoses		Hysterectomy 1990		
49	SM 01	Diagnoses	Diagnoses		Ovarectomy 2010		
50	SM 01	Diagnoses	Diagnoses		Vitamin B 12 anemia		
51	SM 01	Diagnoses	Diagnoses		hypertension		
52	SM 01	Diagnoses	Diagnoses		Arterial hypertension		
53	SM 01	Diagnoses	Diagnoses		Bicuspid aortic valve		

DATA

- 29 Patients
    - 33 common features
    - 199 different features in total

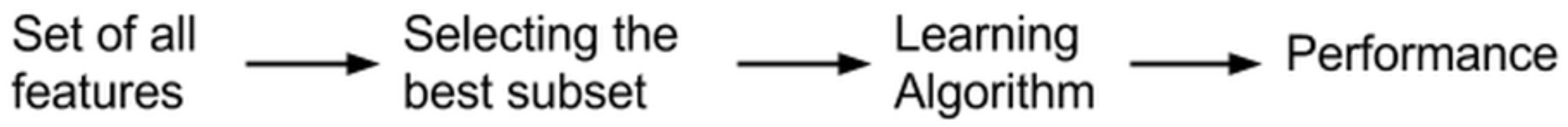
- Feature selection for heterogeneous data

FEATURE SELECTION MODELS

1. Filter
2. Wrapper (DT)
3. Embedded

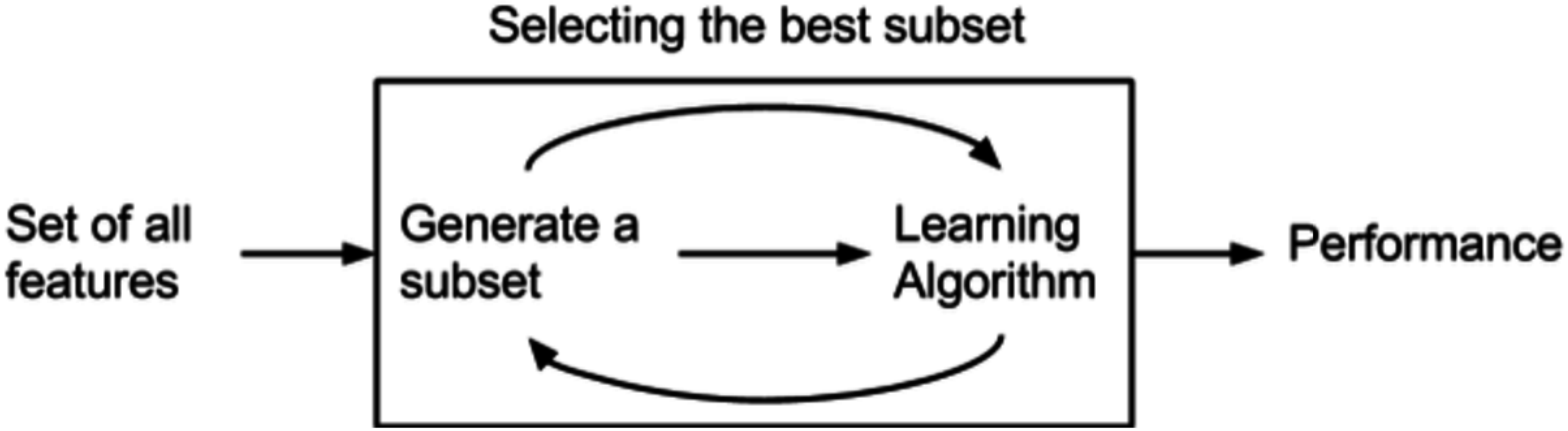
FEATURE SELECTION MODELS - FILTER

- Filters select features based on criteria independent of any supervised learner, performance of filters may not be optimal for a chosen learner
  - Pearson’s Correlation
  - LDA: Linear discriminant analysis
  - ANOVA: Analysis of variance
  - Chi-Square



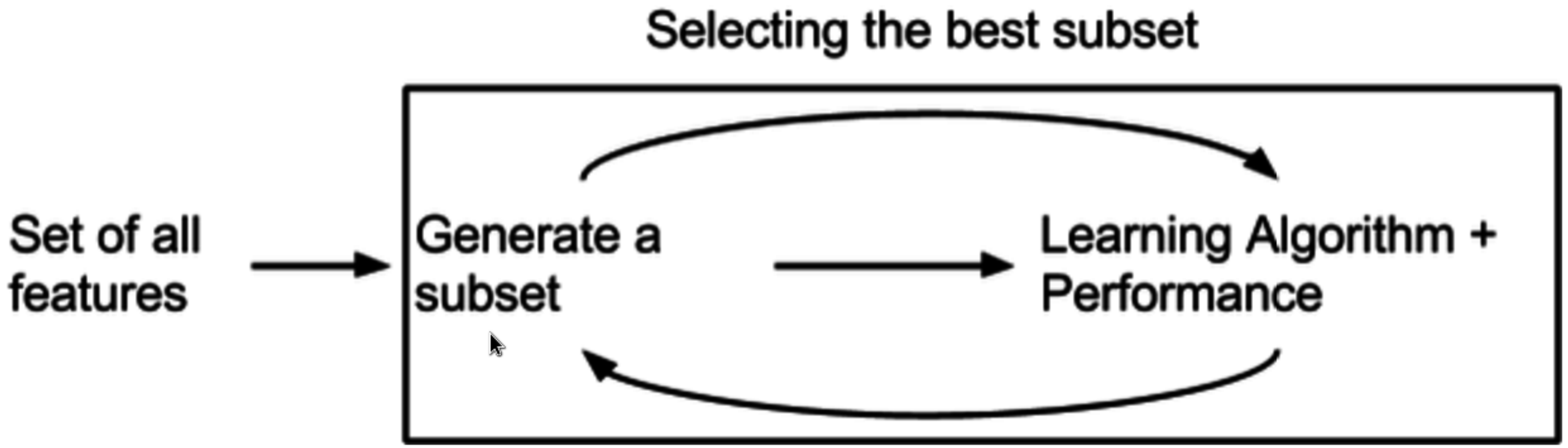
FEATURE SELECTION MODELS - WRAPPER

- Wrappers use learner as black box to evaluate relative usefulness of a feature subset, search the best feature subset for given supervised learner, tend to be computationally expensive
  - Forward Selection
  - Backward Elimination
  - Recursive Feature Elimination



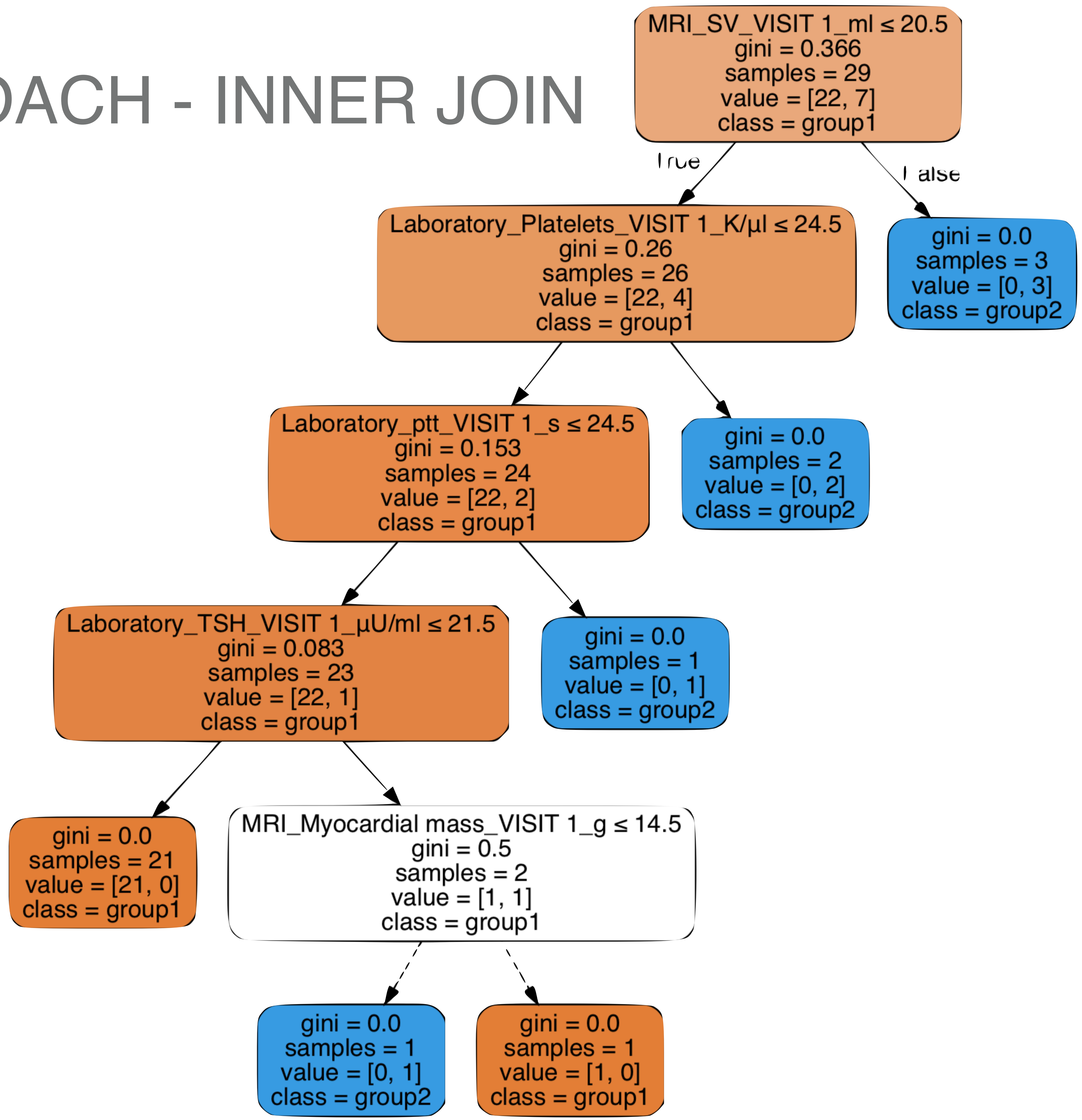
FEATURE SELECTION MODELS - EMBEDDED

- Instead of treating learner as black box, embedded methods select features using the information gained from training a learner
  - Lasso regression (L1 regularization)
  - Ridge regression (L2 regularization)



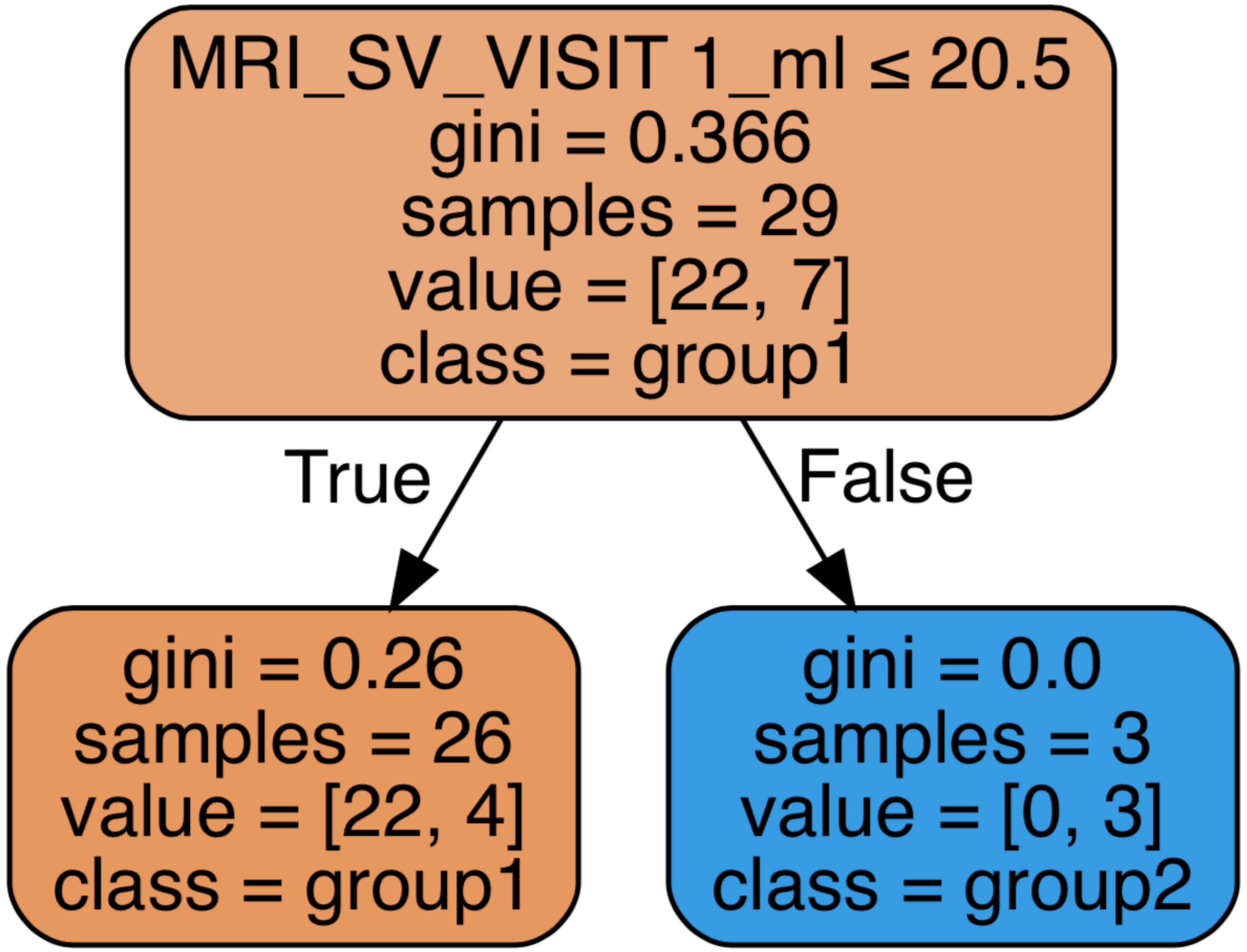
DECISION TREE APPROACH - INNER JOIN

33 CATEGORIES



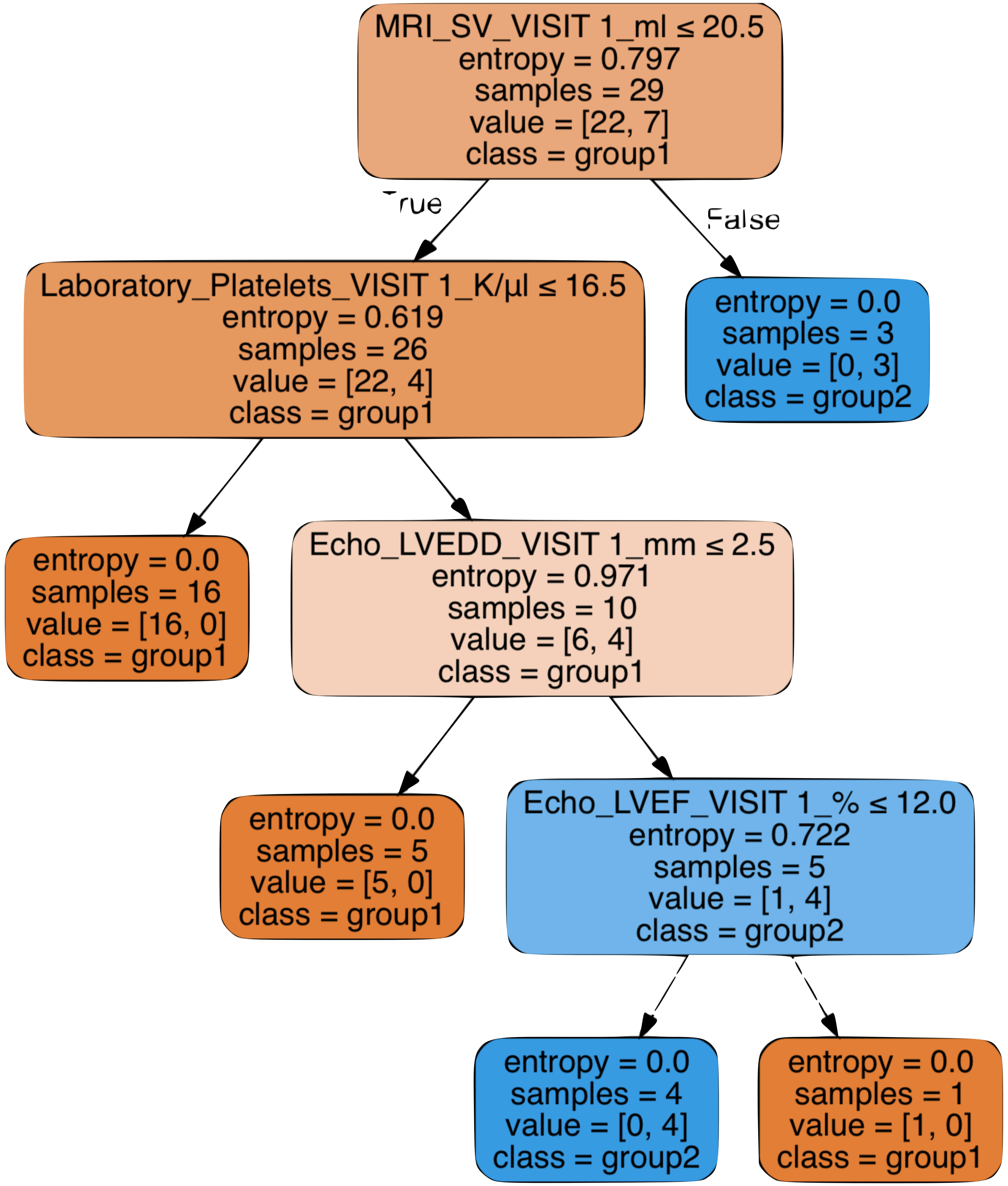
DECISION TREE APPROACH - INNER JOIN & MAX. TREE DEPTH=1

33 CATEGORIES



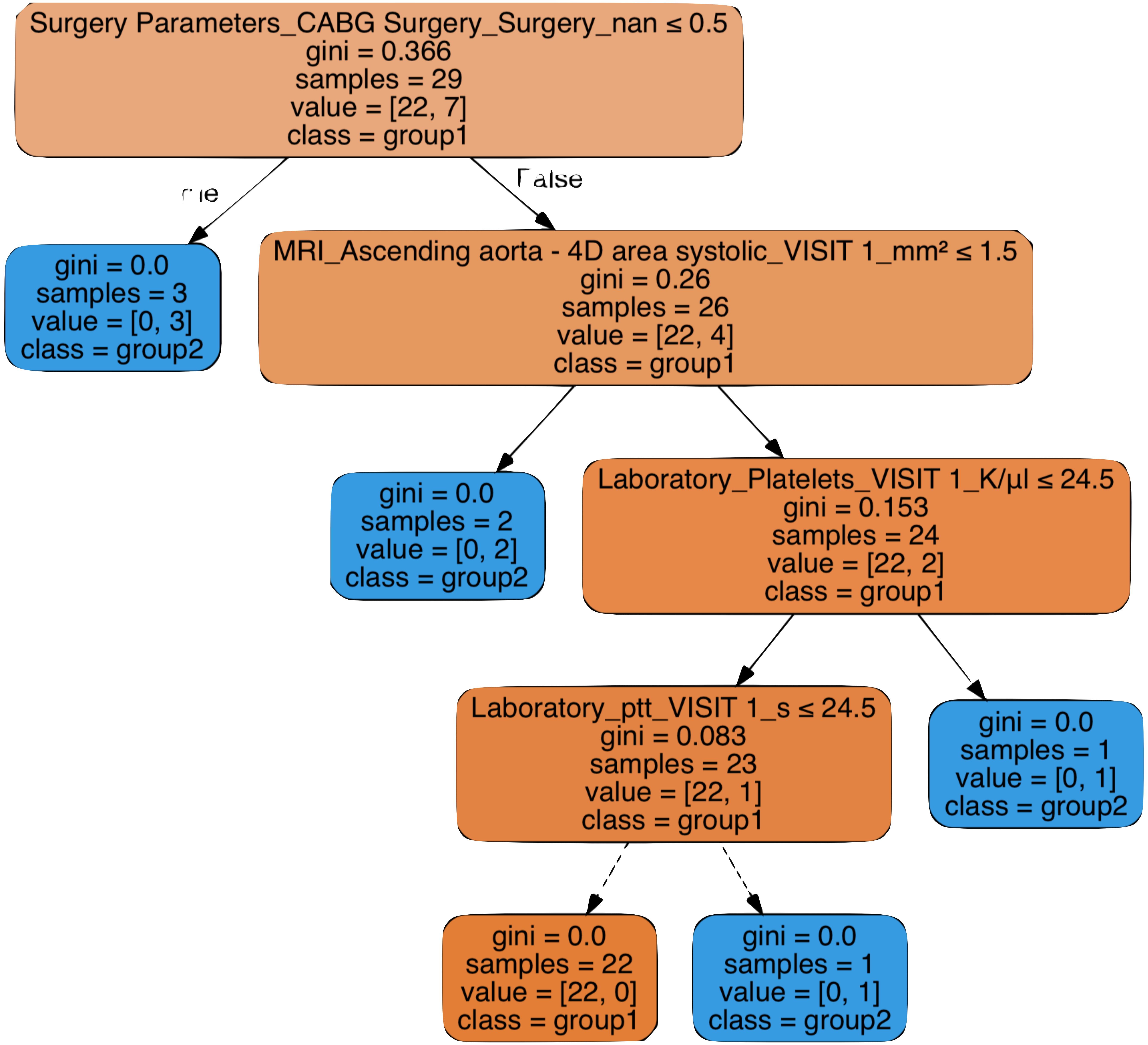
DECISION TREE APPROACH - INNER JOIN & SPLIT-CRITERIA=ENTROPY

33 CATEGORIES



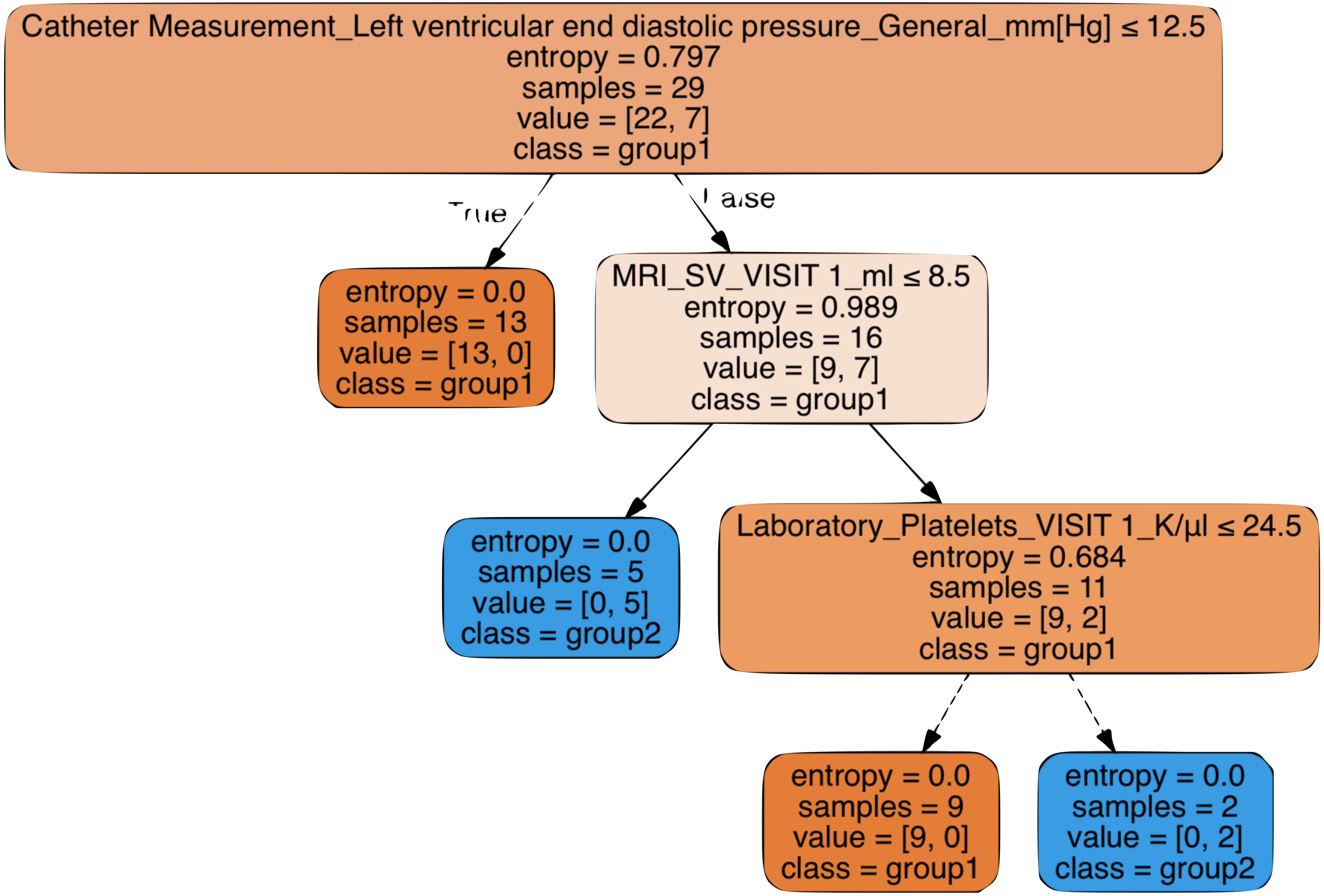
DECISION TREE APPROACH - OUTER JOIN

199 CATEGORIES



DECISION TREE APPROACH - OUTER JOIN & SPLIT_CRITERIA=ENTROPY

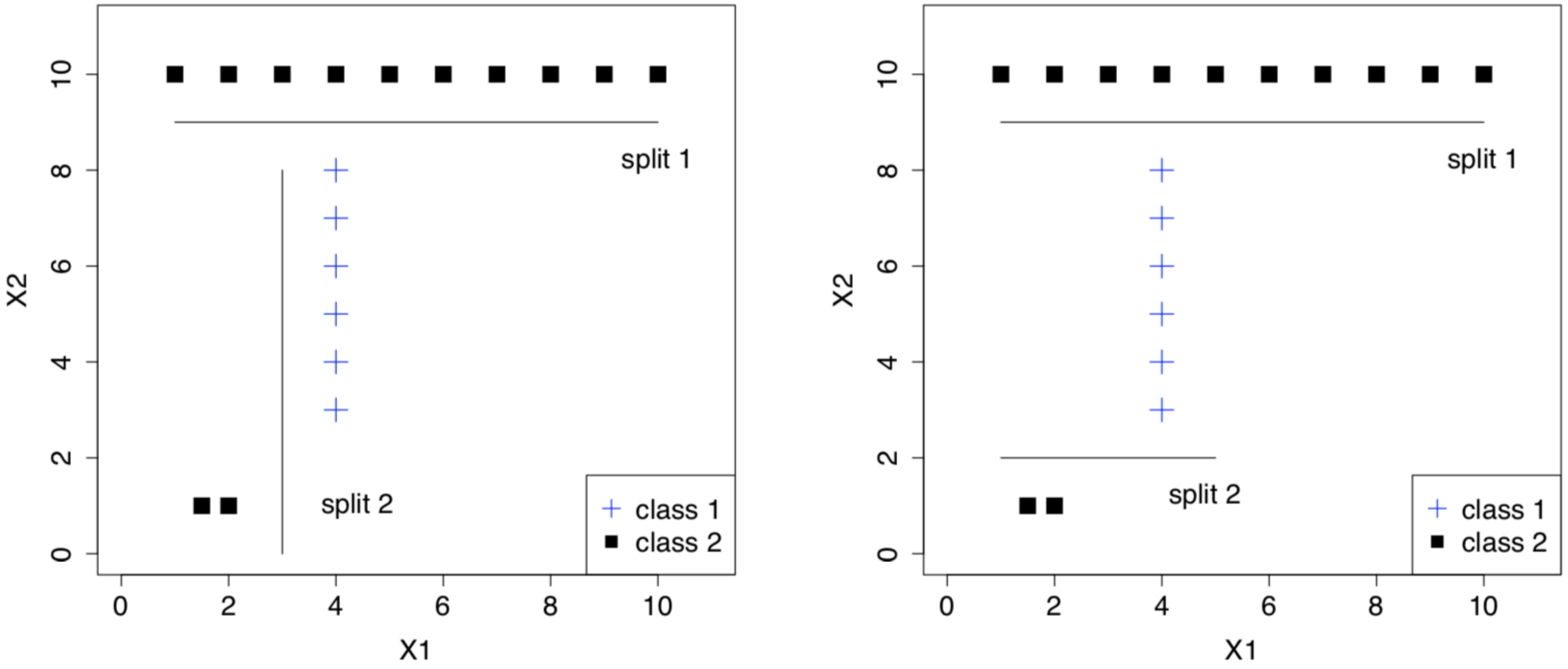
199 CATEGORIES



DECISION TREE APPROACH - STEPS

- Data Preparation
  - 33 common categories in 29 files
  - 199 different categories
  - <u>170</u> categories ignored
- Different imputation strategy
  - mean, median, most frequent value in category, regression

DECISION TREE APPROACH - REDUNDANCY



(a) A decision tree may use both X_1 and X_2 to split. (b) X_2 alone can perfectly separate the two classes.

REGULARIZED TREE MODEL

- Split a decision tree node only if information gain increases significantly
- Lower redundancy
- Higher efficiency
- Apply to regularized bagged random forest

FUTURE WORK

- Implement Regularized Bagged Random Forest (RRF) (https://goo.gl/BqMSMr)
- Implement normal Random Forest classifier for comparison
- Apply other two feature selection models: Filter & Embedded
- Imputation strategy
- Evaluation of results with clinician