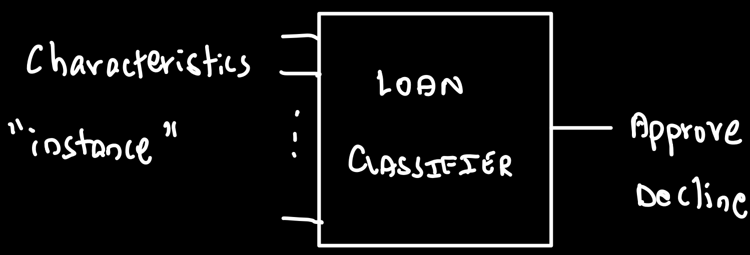
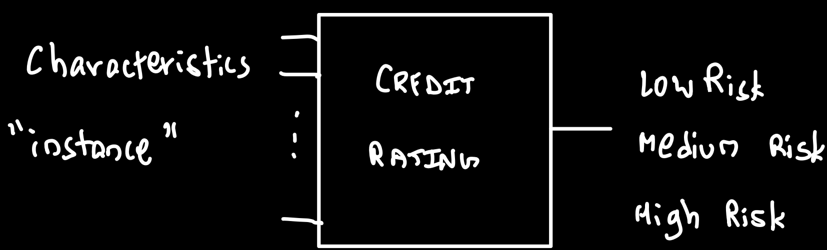
DECISION TREES AND RANDOM FORESTS:

Classifier : type ML Systems





- BAYESIAN NETWORK
  ↳ Classifier.

- Neural Networks Classifier

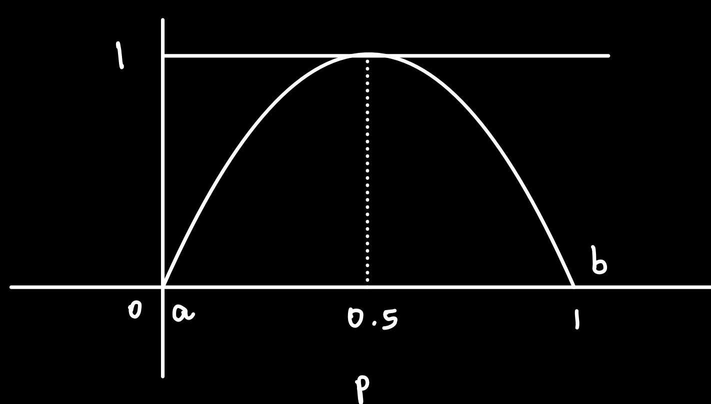
"ENTROPY" : Quantify uncertainty

Entropy:

	Earthquake	Burglary	Alarm
True	0.1	0.2	0.2442
False	0.9	0.8	0.7558
	↓ Most certain		↓ Least certain
ENT(x)	0.469	0.722	0.802

$$ENT(x) = -\sum_{x} Pr(x) \log_2 Pr(x)$$

			a	b
z	z	p	0	1
	$\bar{z}$	1-p	1	0



CONDITIONAL ENTROPY:

$ENT(X)$ $\qquad$ $Y = y$ (observation)

$$ENT(X|y) = -\sum_x Pr(x|y)\log_2 Pr(x|y)$$

$$ENT(X|Y) = \sum_y Pr(y)\, ENT(X|y)$$

$$\boxed{ENT(X|Y) \leq ENT(X)}$$

	$b, \bar{b}$ Burglary	a Burglary | A = true	$\bar{a}$ Burglary | A = false
True	0.2	0.741	0.025
False	0.8	0.259	0.975
ENT	0.722	0.825	0.169
		more uncertain	more certain

$$ENT(\text{Burglary} \mid \text{Alarm}) = ENT(B|a)\,Pr(a) + ENT(B|\bar{a})\,Pr(\bar{a})$$

$$= 0.329 \leq 0.722$$

CLASSIFIERS:

Supervised Learning

Labeled Data

featured attributes]	F_1	F_2	$\ldots,$	F_n	C
e_1					
e_2					
$\vdots$					
e_n					

Decision Tree

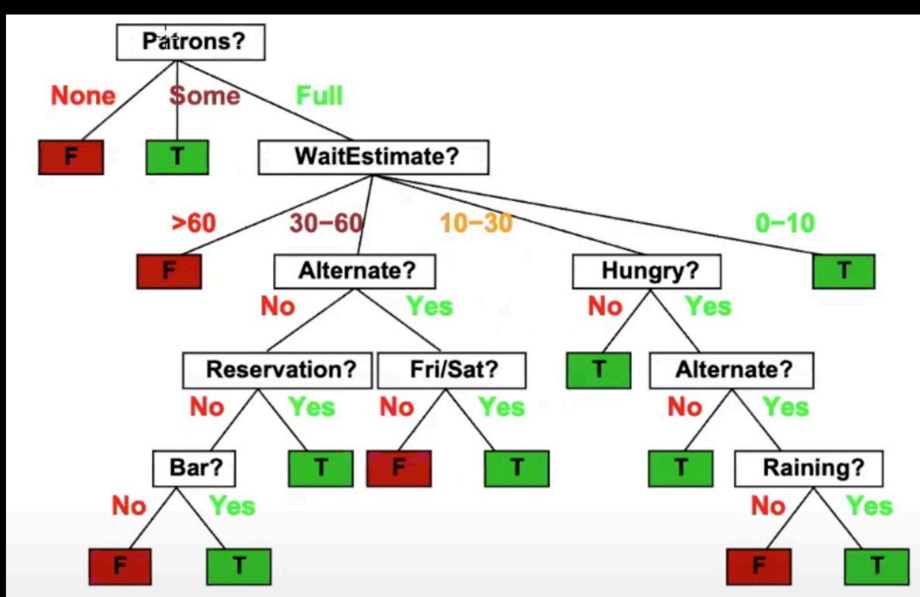
"Random Forest".

Labeled Data:

Example	Attributes										Target
	Alt	*Bar*	*Fri*	*Hun*	*Pat*	*Price*	*Rain*	*Res*	*Type*	*Est*	*WillWait*
X_1	T	F	F	T	Some	$$$	F	T	French	0–10	T
X_2	T	F	F	T	Full	$	F	F	Thai	30–60	F
X_3	F	T	F	F	Some	$	F	F	Burger	0–10	T
X_4	T	F	T	T	Full	$	F	F	Thai	10–30	T
X_5	T	F	T	F	Full	$$$	F	T	French	>60	F
X_6	F	T	F	T	Some	$$	T	T	Italian	0–10	T
X_7	F	T	F	F	None	$	T	F	Burger	0–10	F
X_8	F	F	F	T	Some	$$	T	T	Thai	0–10	T
X_9	F	T	T	F	Full	$	T	F	Burger	>60	F
X_{10}	T	T	T	T	Full	$$$	F	T	Italian	10–30	F
X_{11}	F	F	F	F	None	$	F	F	Thai	0–10	F
X_{12}	T	T	T	T	Full	$	F	F	Burger	30–60	T

12 examples

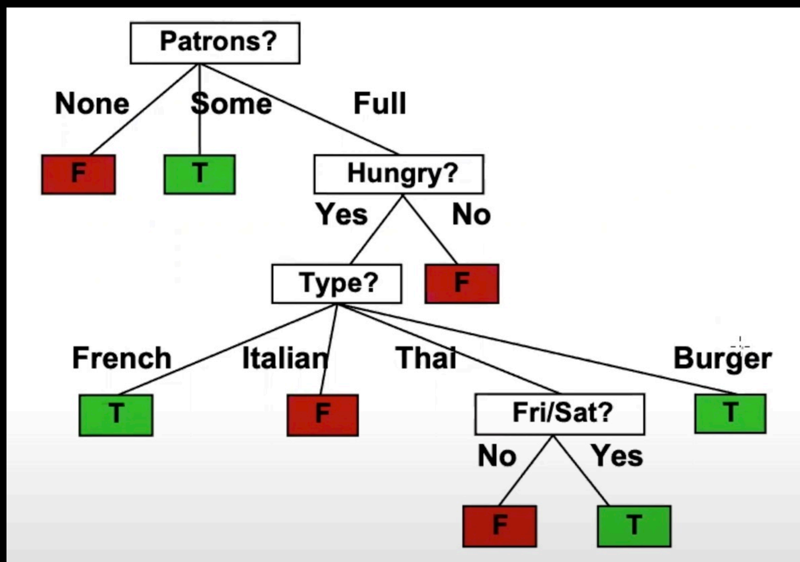
depth



Leaves (Classes)

"Interpretable"

Splitting / Branching



First say we chose Patrons to split

\+ 1, 3, 4, 6, 8, 12

\- 2, 5, 7, 9, 10, 11

Patrons?

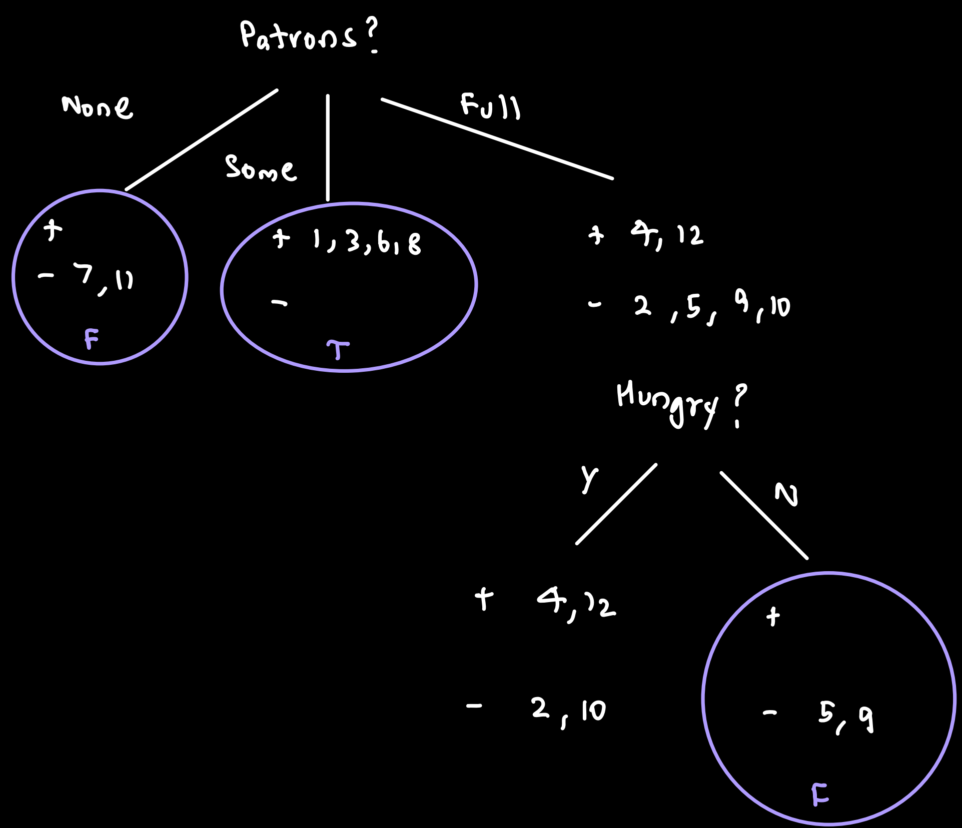
- None
  - \+
  - \- 7, 11
  - F
- Some
  - \+ 1, 3, 6, 8
  - \-
  - T
- Full
  - \+ 4, 12
  - \- 2, 5, 9, 10

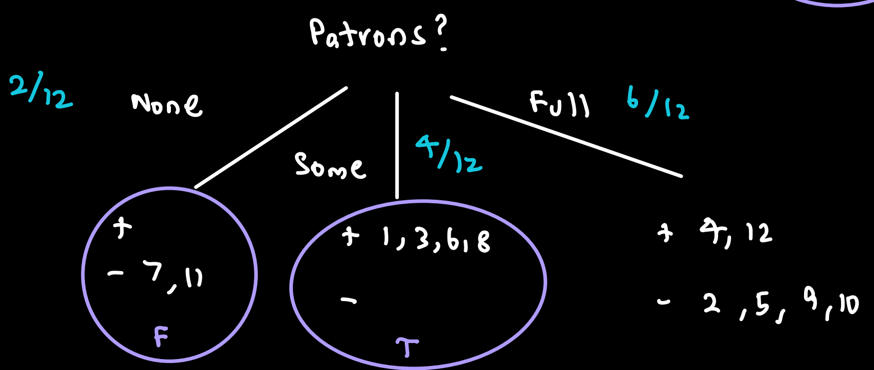
What if we chose Type

Type?

- French
  - \+ 1
  - \- 5
- Italian
  - \+ 6
  - \- 10
- Thai
  - \+ 4, 8
  - \- 2, 11
- Burger
  - \+ 3, 12
  - \- 7, 9

So, which to choose?





ww	
T	0
F	1

Entropy x

ww	
T	1
F	0

Entropy y

ww	
T	$2/6 = 1/3$
F	$4/6 = 2/3$

Entropy z

$$ENT(ww \mid Patrons) = \frac{2}{12}x + \frac{4}{12}y + \frac{6}{12}z$$

+50
-50
A

yes — no

+30
-40

+20
-10

M = low
high

M	
low	30/70
high	40/70

$\underbrace{\qquad}_{0.985}$

M	
low	20/30
high	10/30

$\underbrace{\qquad}_{0.918}$

$$ENT(M \mid A) = 0.7 \times 0.985 + 0.3 \times 0.918$$

$$= 0.965$$

+50
-50
B

low — mid — high

+20
-15

+20
-15

+10
-20

$\underbrace{\qquad\qquad\qquad}_{\text{score}}$

if no examples left to split, majority of parents' answers is returned.

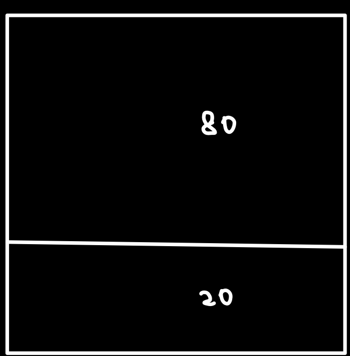
function DECISION-TREE-LEARNING(*examples*, *attributes*, *parent_examples*) returns a tree
  if *examples* is empty then return PLURALITY-VALUE(*parent_examples*)
  else if all *examples* have the same classification then return the classification
  else if *attributes* is empty then return PLURALITY-VALUE(*examples*)
  else
    $A \leftarrow \text{argmax}_{a \in attributes}$ IMPORTANCE(a, *examples*) → Conditional entropy
    tree ← a new decision tree with root test A
    for each value v_k of A do
      exs ← $\{ e : e \in examples$ and $e.A = v_k \}$
      subtree ← DECISION-TREE-LEARNING(*exs*, *attributes* − A, *examples*)
      add a branch to *tree* with label ($A = v_k$) and subtree *subtree*
    return *tree*

if no attribute left to split just return majority out of the examples left.

PLURALITY-VALUE → Majority

TO COMPARE

100



training dataset

testing dataset

NAIVE BAYES CLASSIFIER

Decision Tree

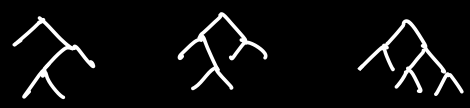
Divide 80:20 - 5 time. [CROSS VALIDATION]

→ Accuracy

Take average accuracy

RANDOM FORESTS:

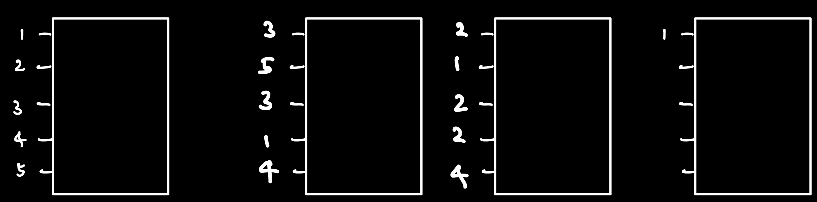
Build bunch of trees.



Instance yes no no => no

ensemble learning methods

BOOTSTRAPPED DATASET



2 3, 5

out-of-bag example

$n = [10 \text{ attributes}]$

If $m = 3$, then at any point we choose 3 attributes to choose from.

Say m can be $\sqrt{n}$

A

Evaluate accuracy on Out-of-bag examples.

1. Ensemble (Majority Voting)

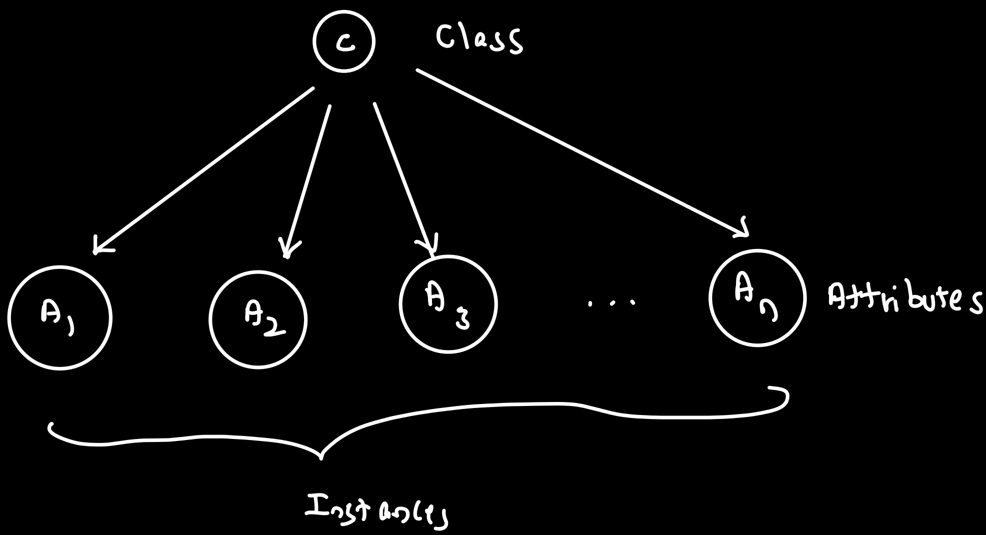
2. Bootstrapped Dataset
   (Randomly pick examples to create dataset)

3. Choose Random Subset of attributes while deciding
   next attribute.

Effectiveness:

Random Forest < Bayesian Network < Neural Network.

BAYESIAN NETWORK CLASSIFIERS:

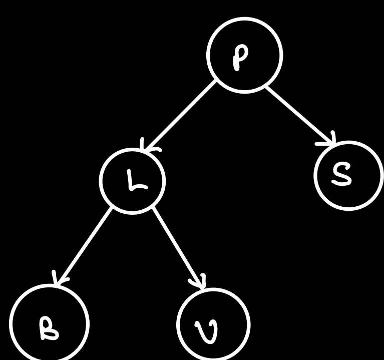


$Pr(c \mid a_1 \ldots a_n)$ $\qquad$ $Pr(\bar{c} \mid a_1 \ldots a_n)$

Threshold

$c \; : \; Pr(c \mid a_1 \ldots a_n) \geq T$

$\bar{c} \; : \; Pr(c \mid a_1 \ldots a_n) < T$



$T = 90\%$

Instance

S	B	U
+	-	-

$Pr(P \mid S = +, B = -, U = -) \geq 90\%$

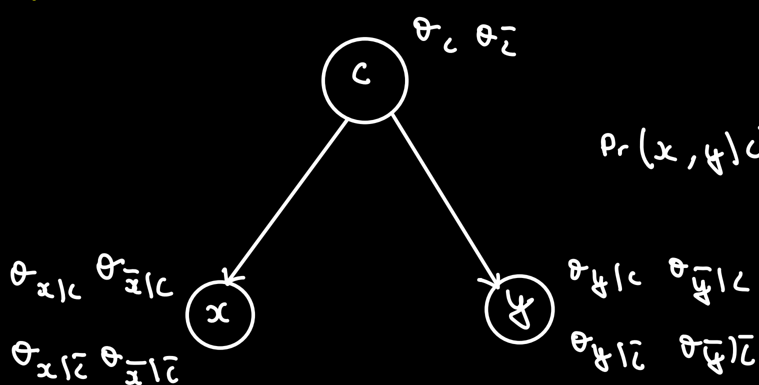
not pregnant $\qquad$ pregnant

$O(n \cdot d^{w})$

trees $\rightarrow w = 1$

Naive Bayes Classifier is a tree

NAIVE BAYES CLASSIFIER:



$$Pr(x, y | c) = Pr(x|c) \, Pr(y|c)$$

$$Pr(c | x, y) = \frac{Pr(x, y | c) \, Pr(c)}{Pr(x, y)}$$ By Bayes Rule

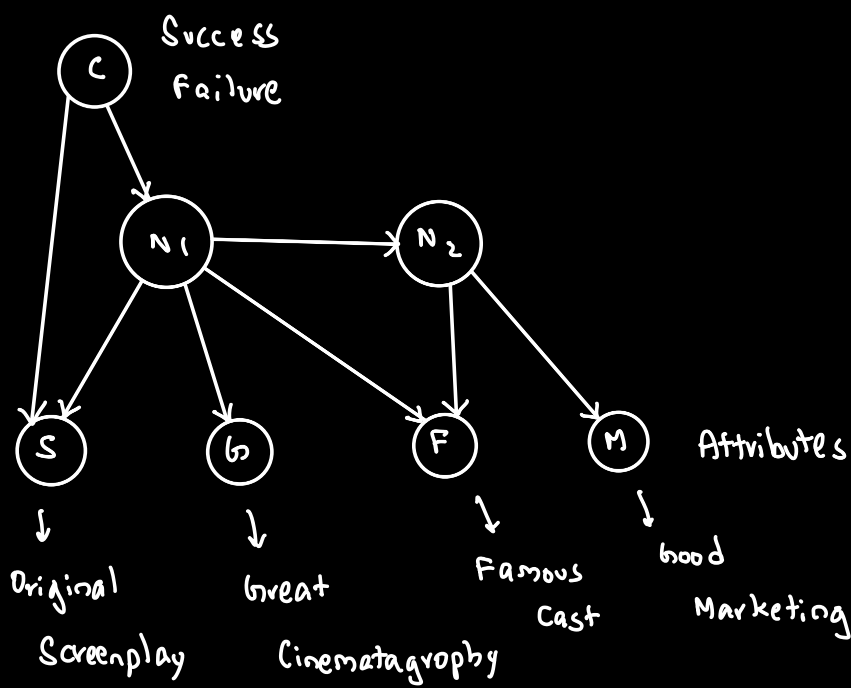
$$= \frac{Pr(x|c) \, Pr(y|c) \, Pr(c)}{Pr(x, y)}$$

$\hookrightarrow Pr(x, y | c) \, Pr(c) +$

$Pr(x, y | \bar{c}) \, Pr(\bar{c})$

So all the values are directly available from the CPT!

EXPLAINABLE AI:



Numeric

↳ Compiles into a circuit with same input-output behavior.

S	G	F	M	C
−	−	−	−	−
−	−	−	+	−
−	−	+	+	+
⋮				⋮
+	+	+	+	+

16

Symbolic

We can represent efficiently using Boolean Circuits.

Tractable Circuits → NNF Circuits

$S=1, G=0, F=1, M=1 \Rightarrow$ Success

Why Success?

Possible Answer: Based on $F=1, M=1$.

This is PI- explanation.

2 Steps:

→ Compile ML system into Symbolic Representation that capture I/O behavior.

→ Use Symbolic Reasoning to reason about the behavior.