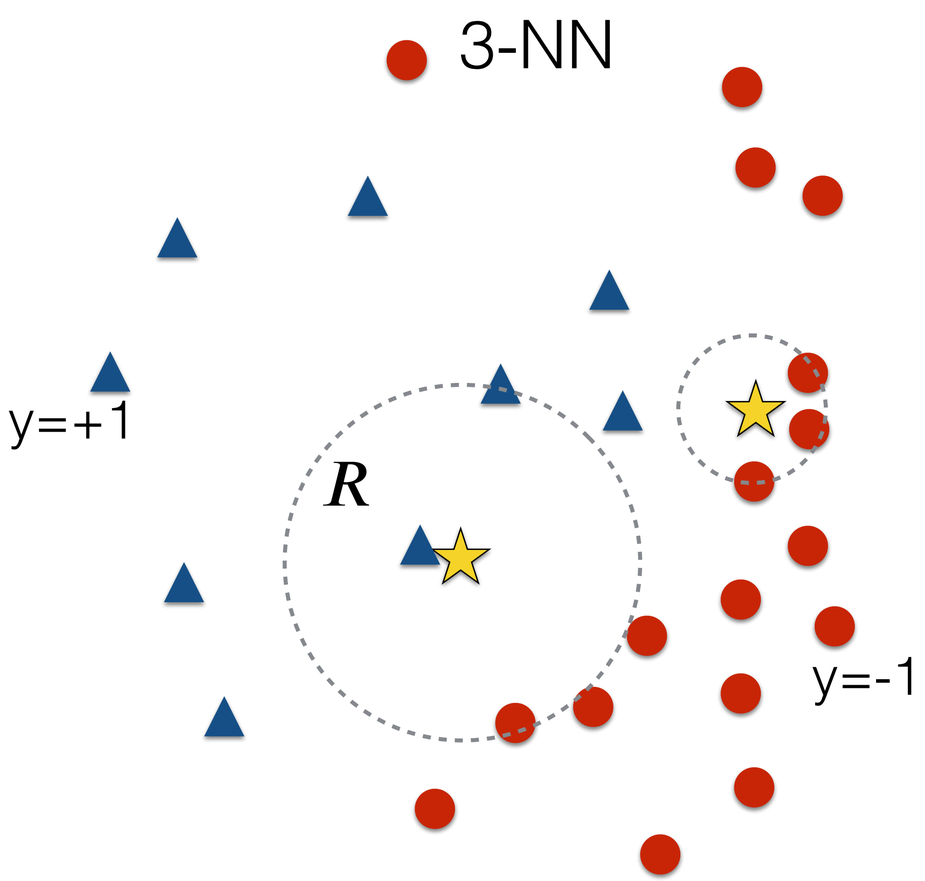
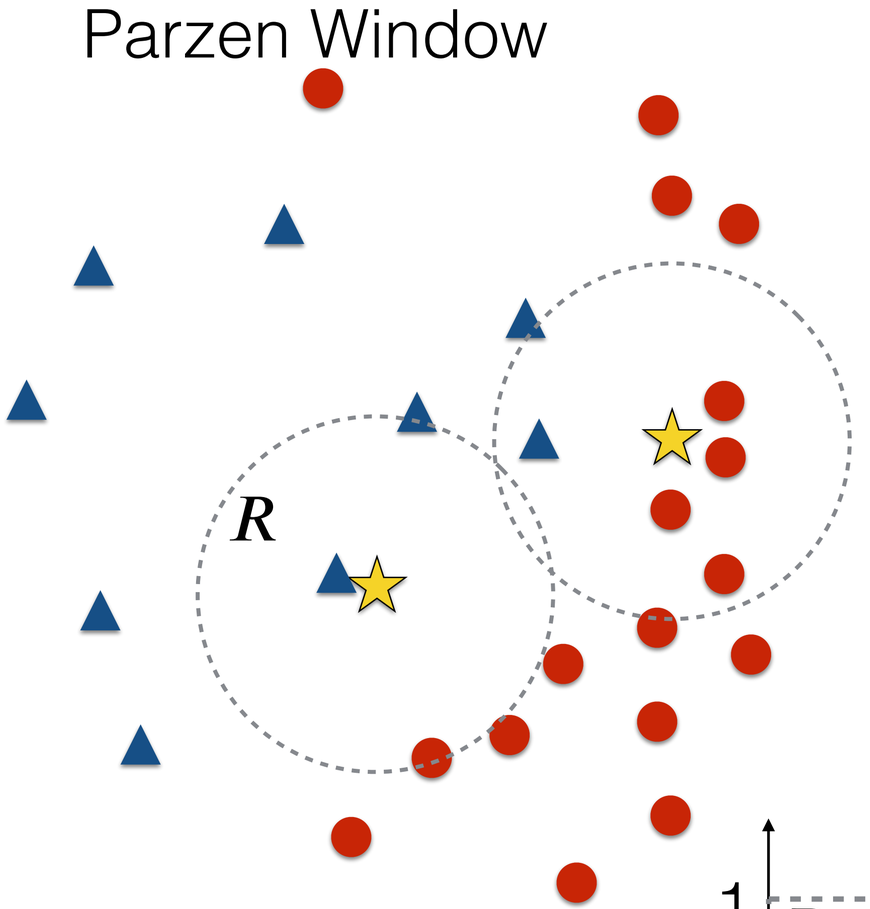
Statistical Methods in AI (CSE 471)
Lecture5: Decision Trees

Vineet Gandhi
Centre for Visual Information Technology (CVIT)

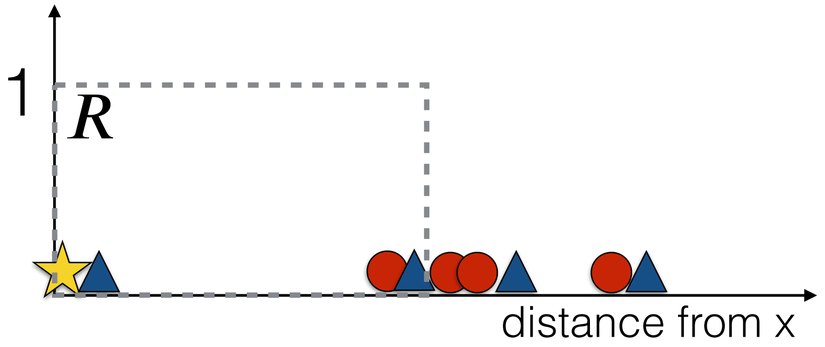
kNN, Parzen Windows and Kernels



$$h(x) = \text{sgn}[\sum_i y_i \cdot 1_{||x_i - x|| \leq R}]$$



$$h(x) = \text{sgn}[\sum_{i:x_i \in R(x)} y_i]$$



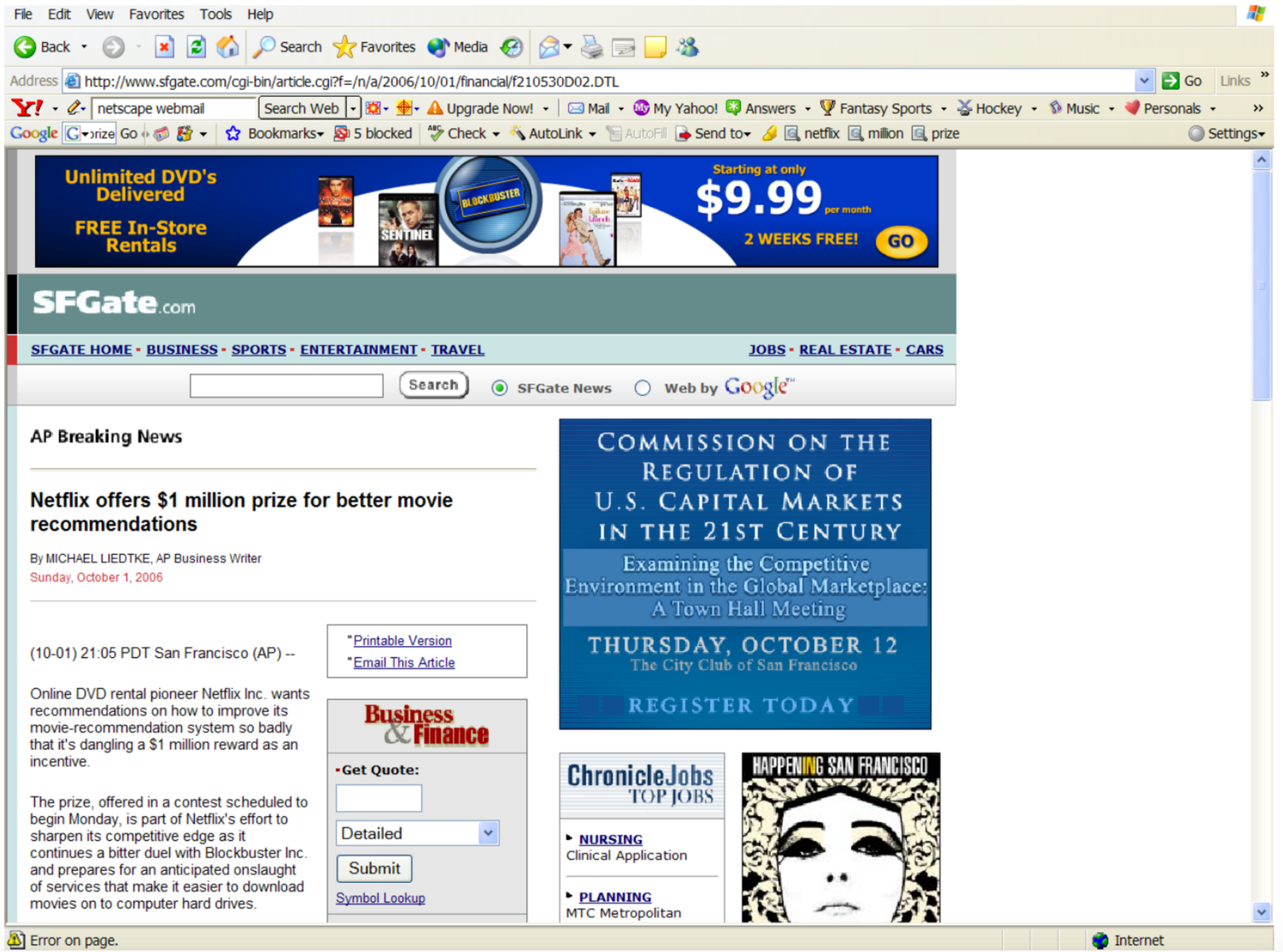
slide credit: Victor Lavrenko

Decision Trees

- One of the most intuitive classifiers
- Easy to understand and construct
- Surprisingly, also works very (very) well

\

Decision Trees



Decision Trees

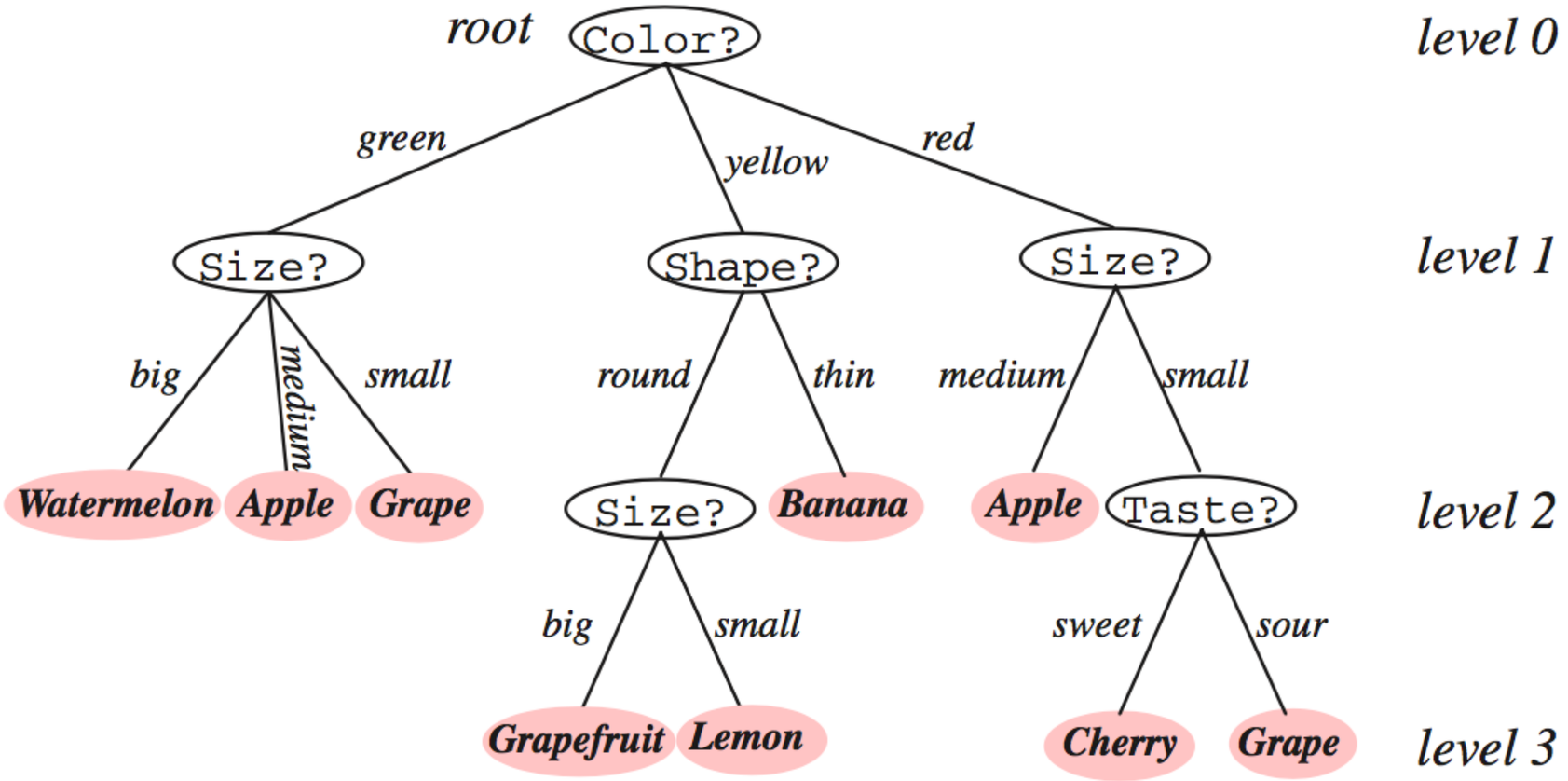
Attributes (features)

Label

Movie	Type	Length	Director	Famous actors	Liked?
m1	Comedy	Short	Adamson	No	Yes
m2	Animated	Short	Lasseter	No	No
m3	Drama	Medium	Adamson	No	Yes
m4	animated	long	Lasseter	Yes	No
m5	Comedy	Long	Lasseter	Yes	No
m6	Drama	Medium	Singer	Yes	Yes
m7	animated	Short	Singer	No	Yes
m8	Comedy	Long	Adamson	Yes	Yes
m9	Drama	Medium	Lasseter	No	Yes

100,480,507 ratings that 480,189 users gave to 17,770 movies

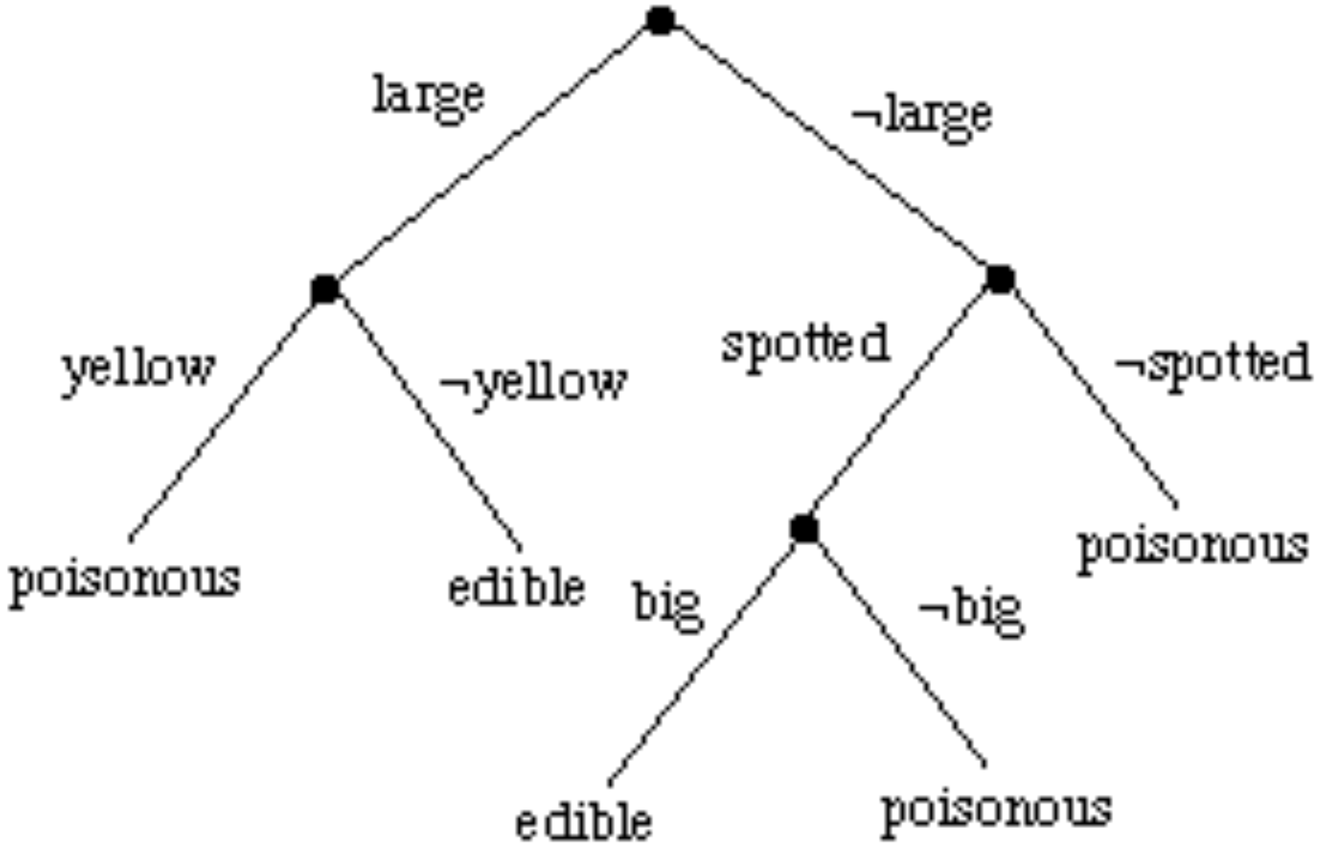
Decision Trees



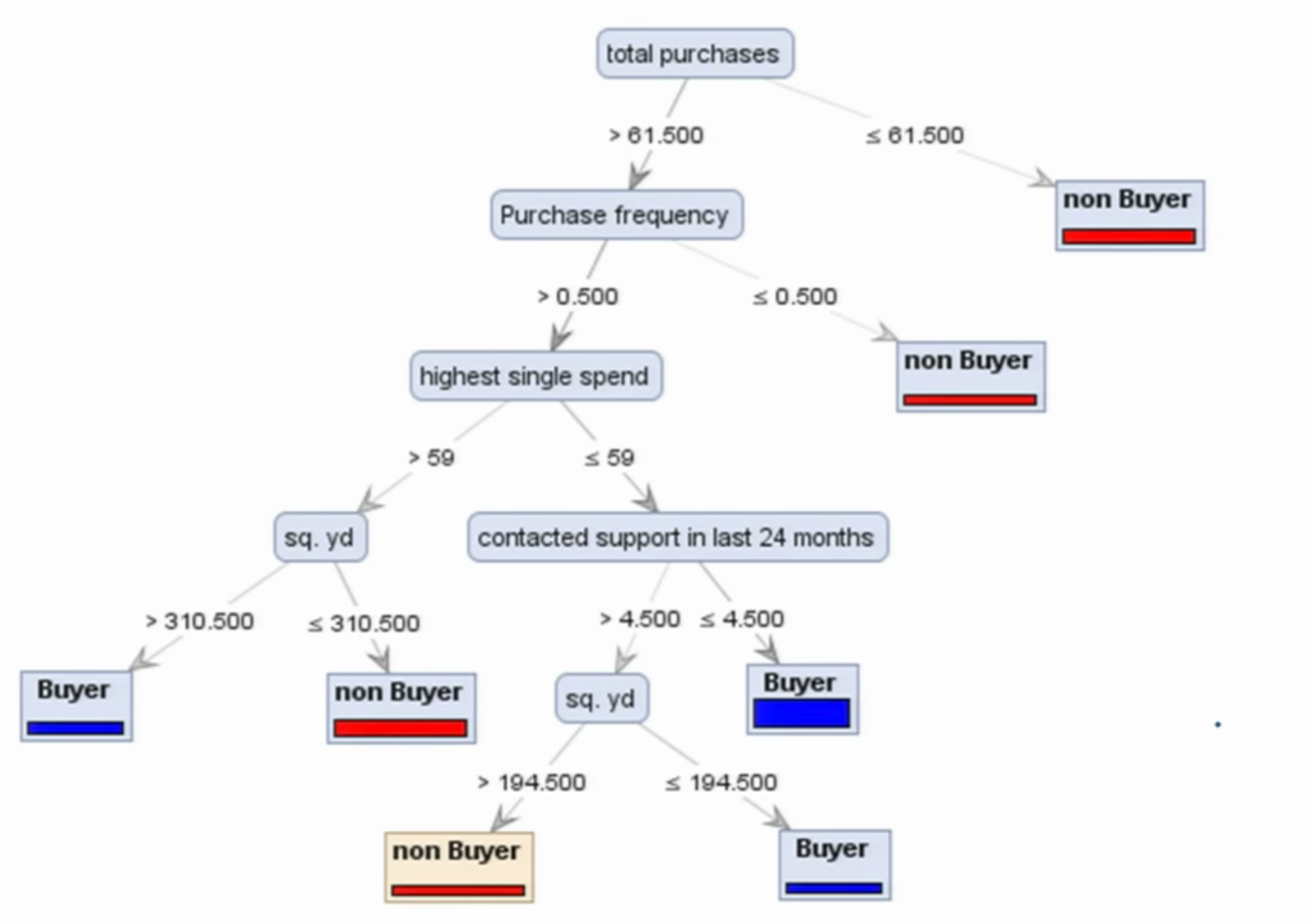
fruit described using 4 tuple: {color, shape, taste, size}

source: “Pattern Classification, Duda”, Hart and Stork

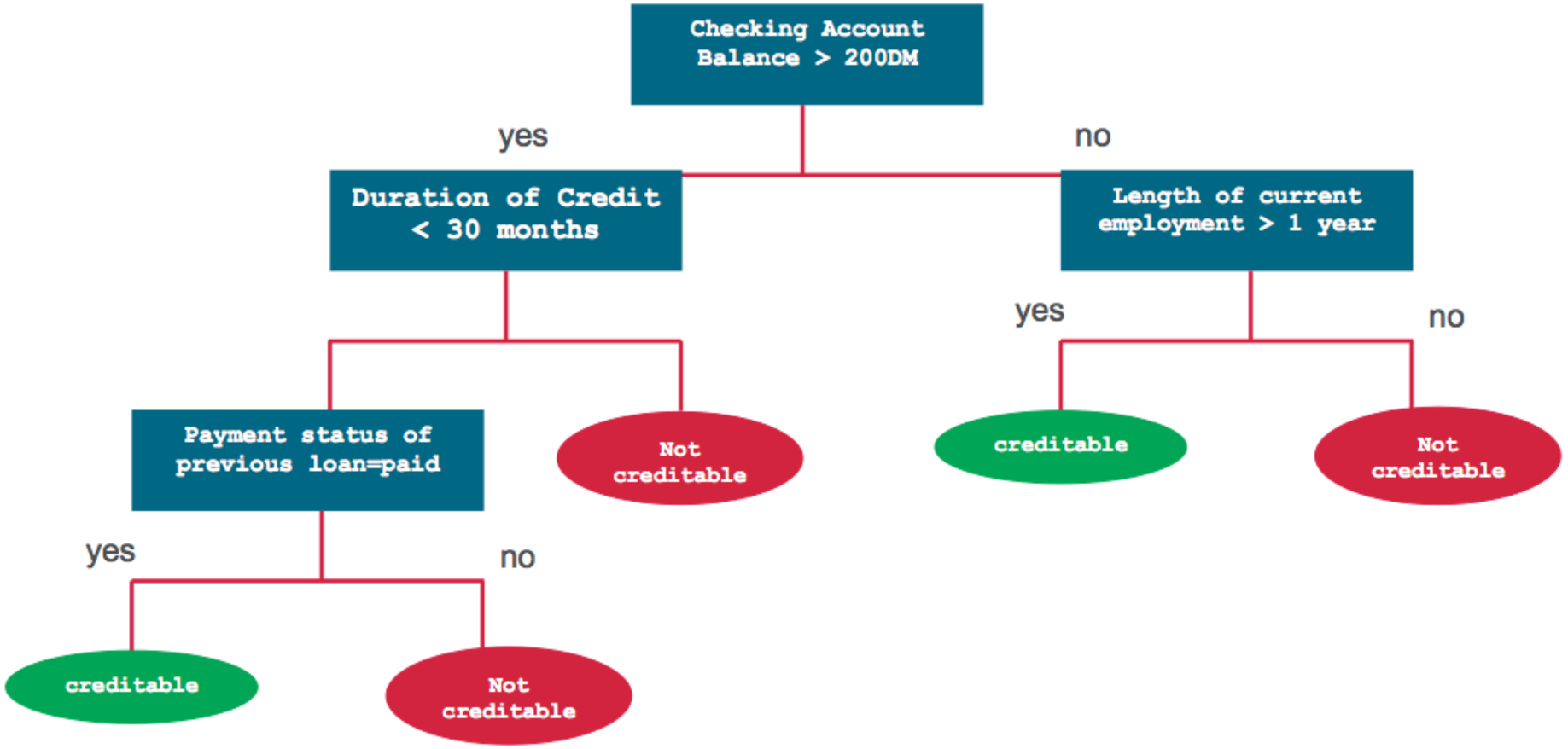
Edible Mushroom



Identify buyers



Credit approval



Vampire or not

Vampire?	Shadow?	Garlic?	Complexion?	Accent?
No	?	Yes	Pale	None
No	Yes	Yes	Ruddy	None
Yes	?	No	Ruddy	None
Yes	No	No	Average	Heavy
Yes	?	No	Average	Odd
No	Yes	No	Pale	Heavy
No	?	Yes	Ruddy	Odd
No	Yes	No	Average	Heavy

dataset credit: “Artificial Intelligence”, Patrick Winston

Vampire or not

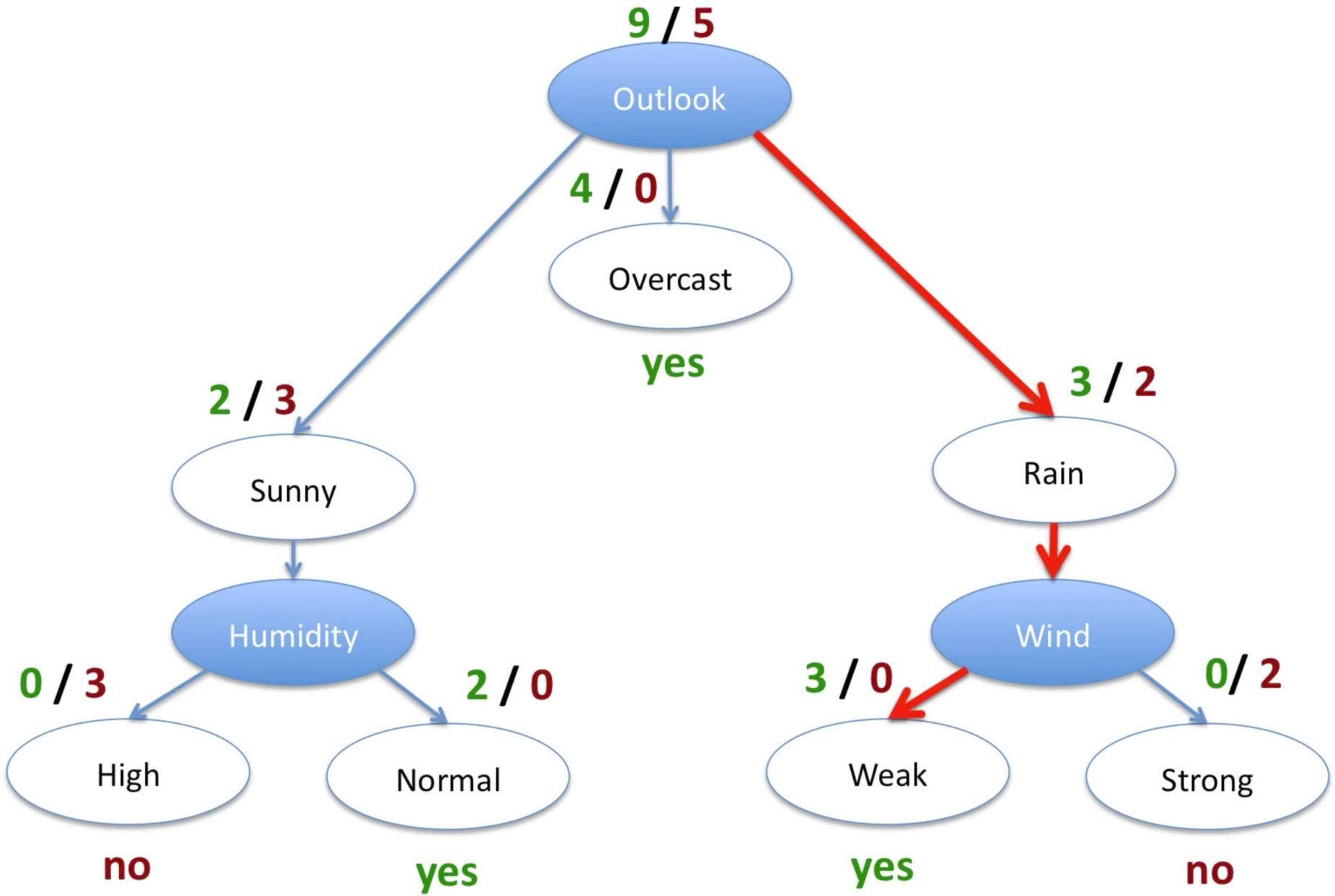
Vampire?	Shadow?	Garlic?	Complexion?	Accent?
No	?	Yes	Pale	None
No	Yes	Yes	Ruddy	None
Yes	?	No	Ruddy	None
Yes	No	No	Average	Heavy
Yes	?	No	Average	Odd
No	Yes	No	Pale	Heavy
No	?	Yes	Ruddy	Odd
No	Yes	No	Average	Heavy

dataset credit: “Artificial Intelligence”, Patrick Winston

Another example

Training examples: **9 yes / 5 no**

Day	Outlook	Humidity	Wind	Play
D1	Sunny	High	Weak	No
D2	Sunny	High	Strong	No
D3	Overcast	High	Weak	Yes
D4	Rain	High	Weak	Yes
D5	Rain	Normal	Weak	Yes
D6	Rain	Normal	Strong	No
D7	Overcast	Normal	Strong	Yes
D8	Sunny	High	Weak	No
D9	Sunny	Normal	Weak	Yes
D10	Rain	Normal	Weak	Yes
D11	Sunny	Normal	Strong	Yes
D12	Overcast	High	Strong	Yes
D13	Overcast	Normal	Weak	Yes
D14	Rain	High	Strong	No

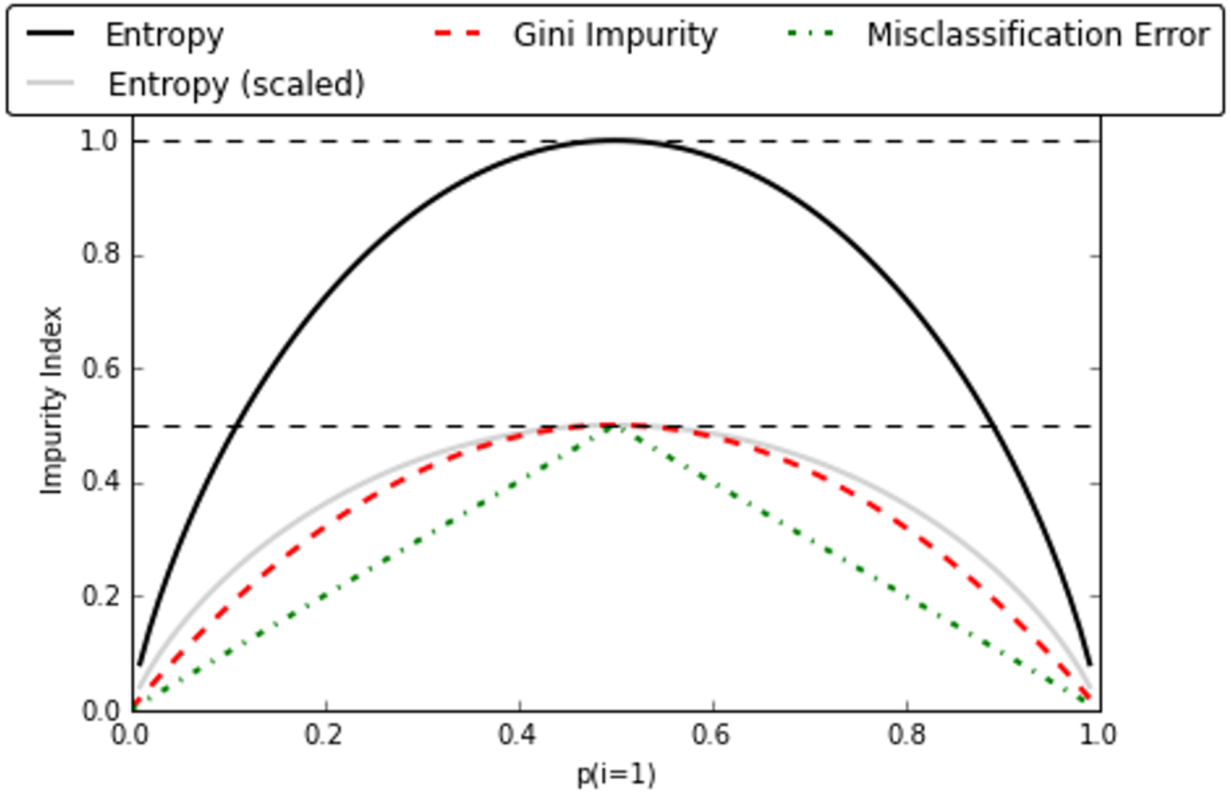

9 / 5
Outlook
4 / 0
Overcast
yes
2 / 3
Sunny
3 / 2
Rain
Humidity
Wind
0 / 3
2 / 0
3 / 0
0/ 2
High
Normal
Weak
Strong
no
yes
yes
no

Impurity functions

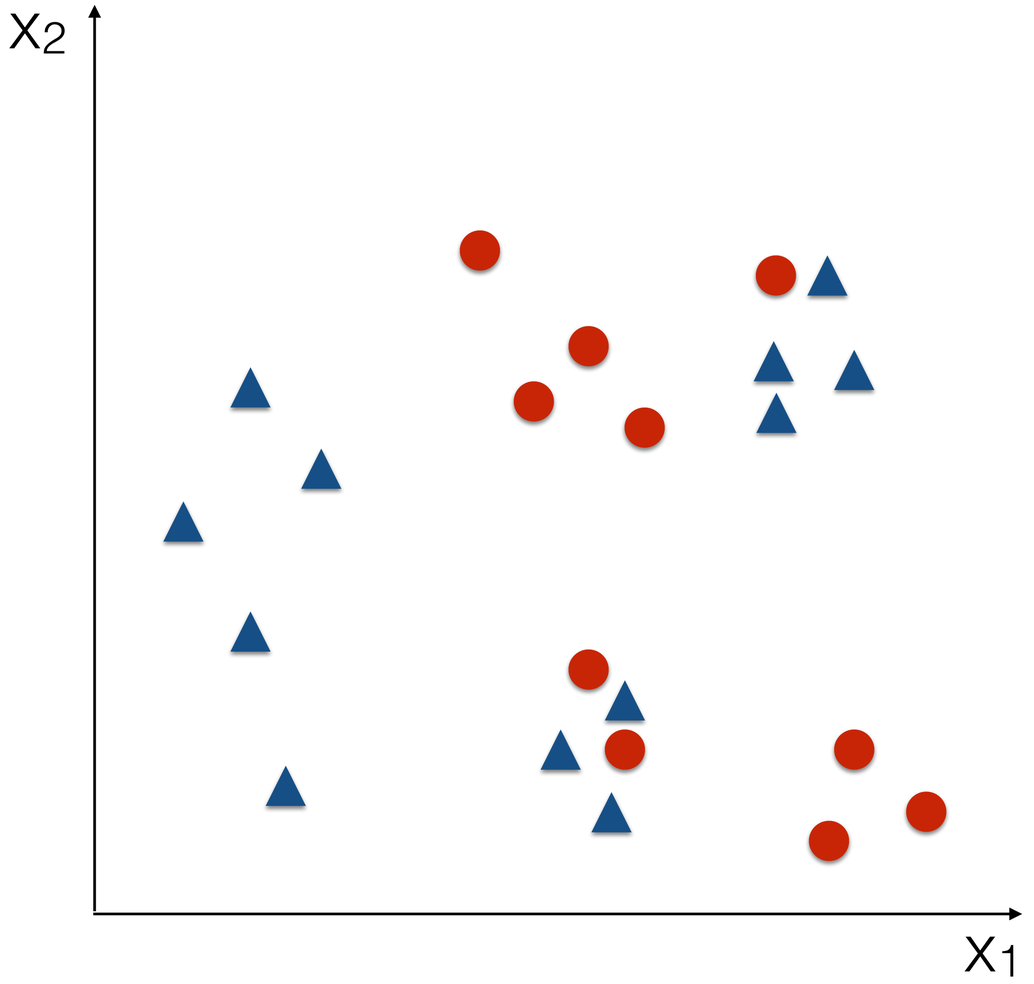
Entropy: $i(V) = -\big(q \log q + (1-q)\log(1-q)\big)$

Gini index: $i(V) = 2q(1-q)$

Misclassification rate: $i(V) = \min(q, 1-q)$

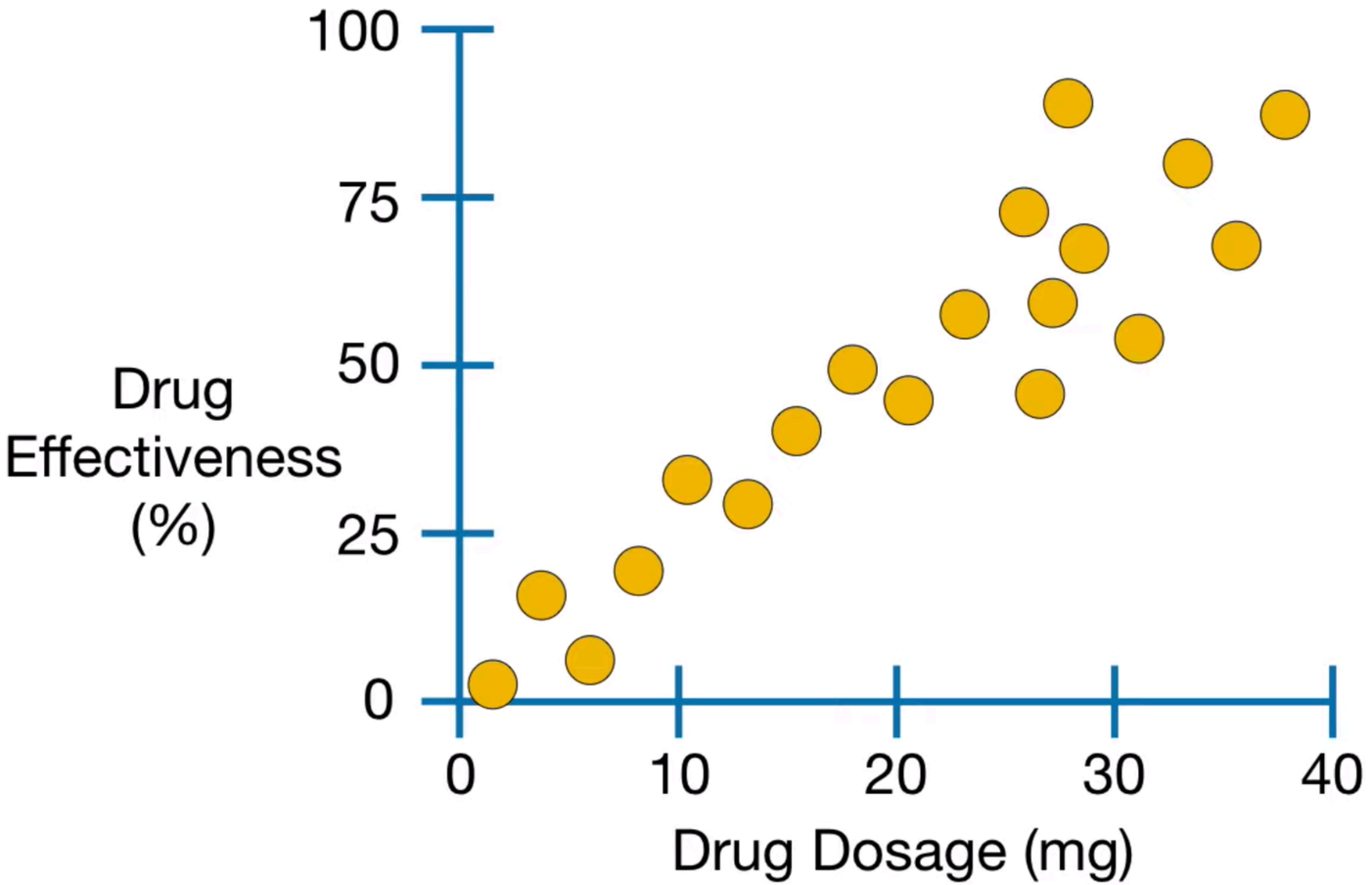


Numeric data?

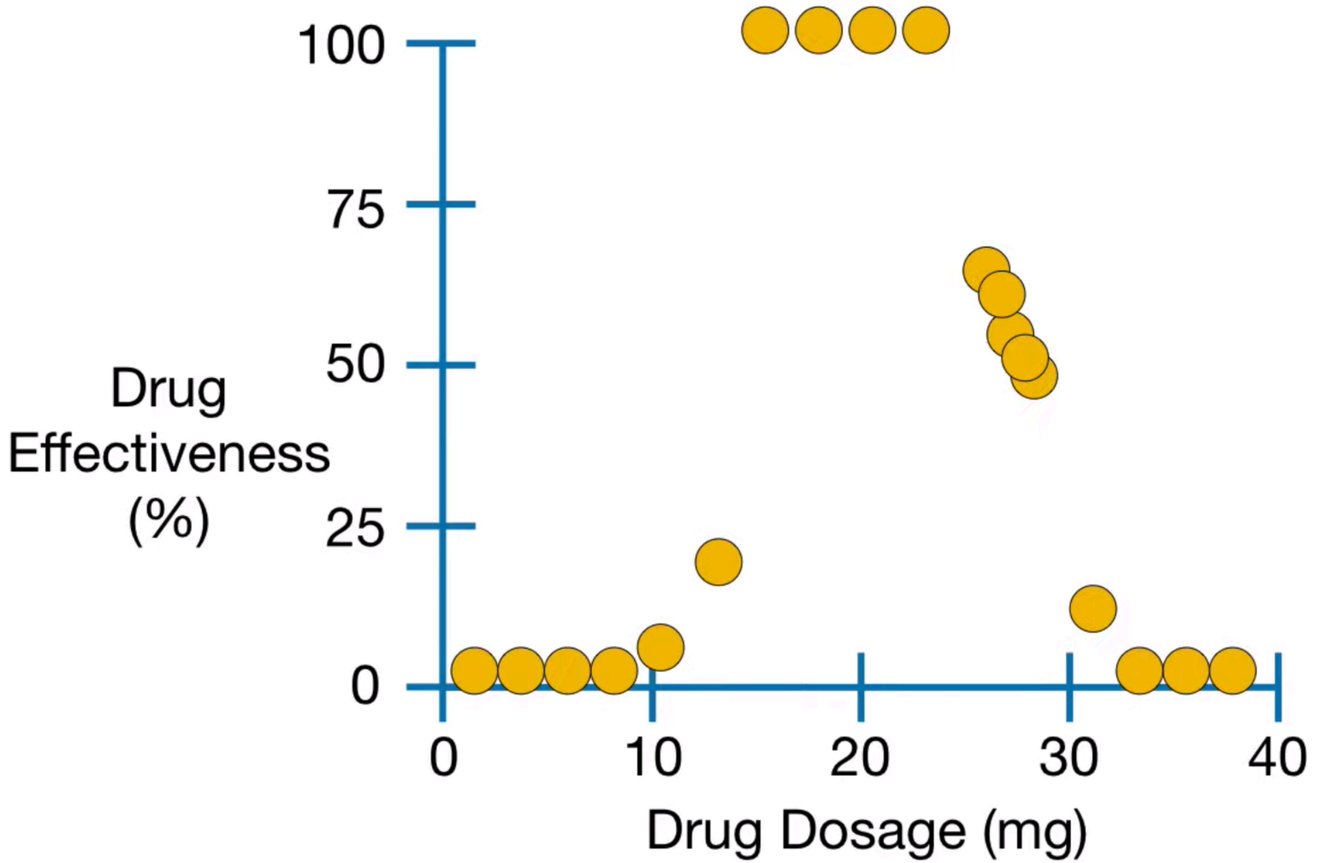


Regression

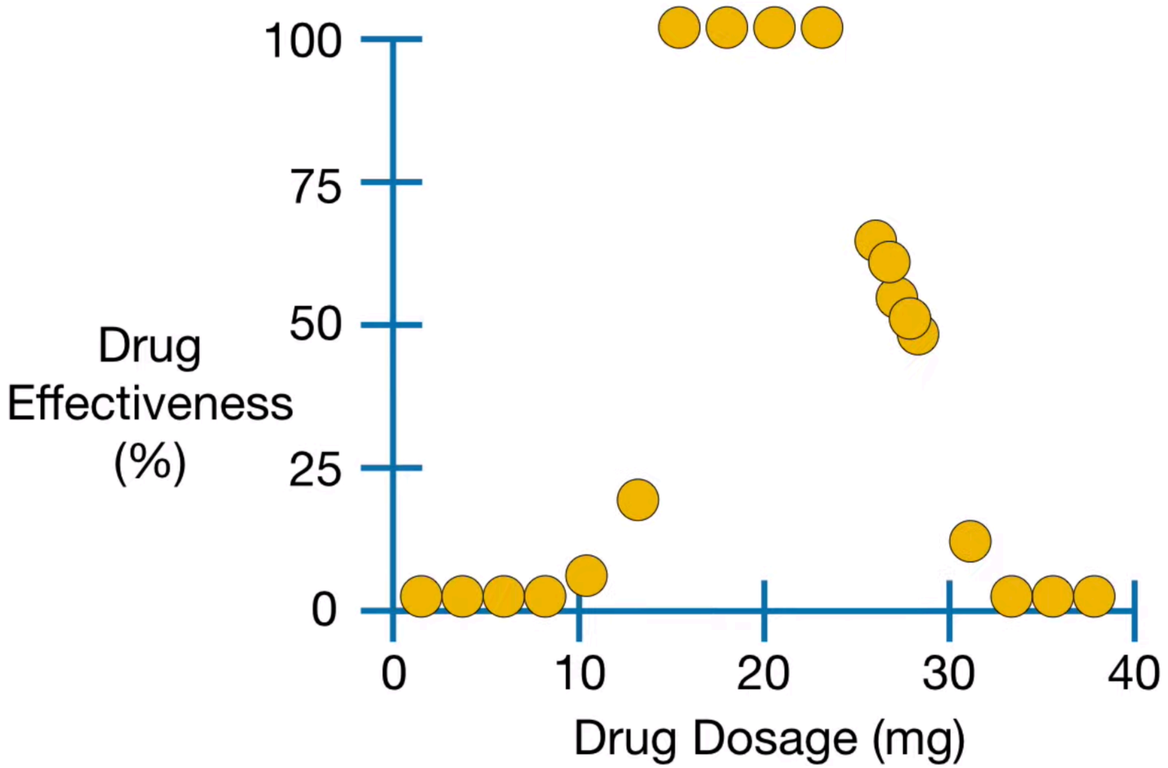
- Drug effectiveness vs dosage

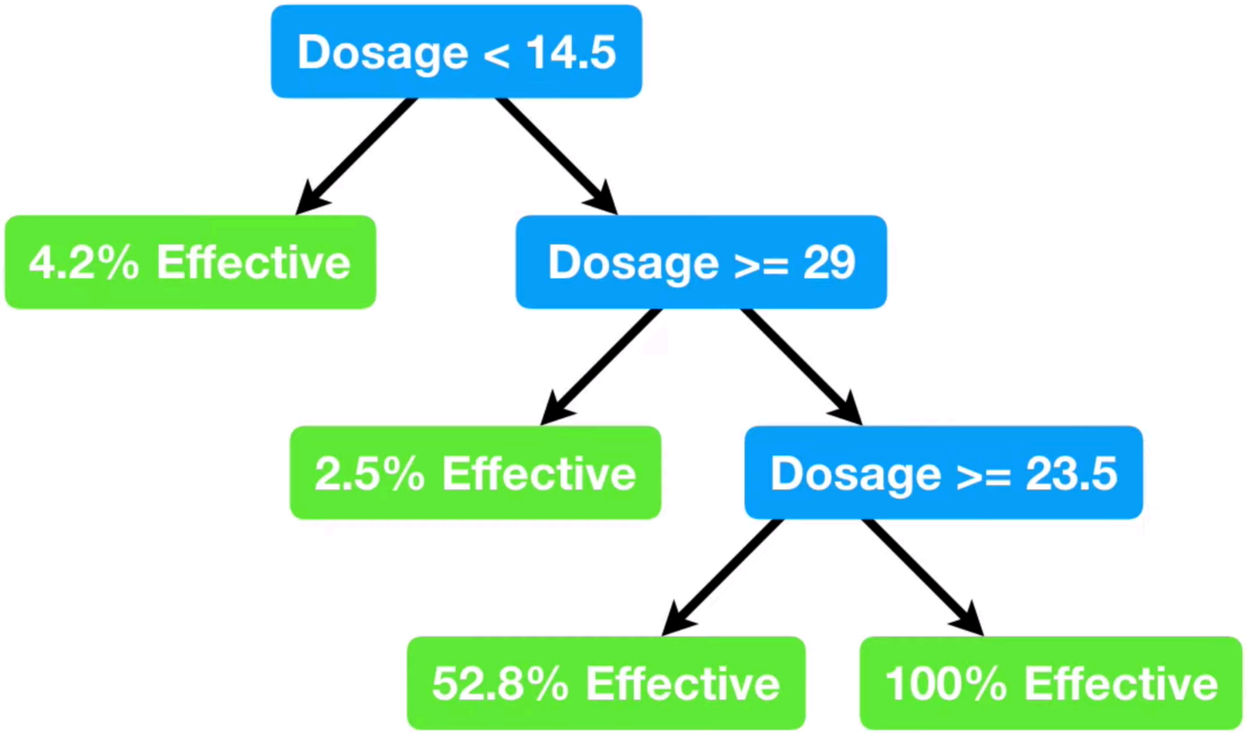


Regression



Regression

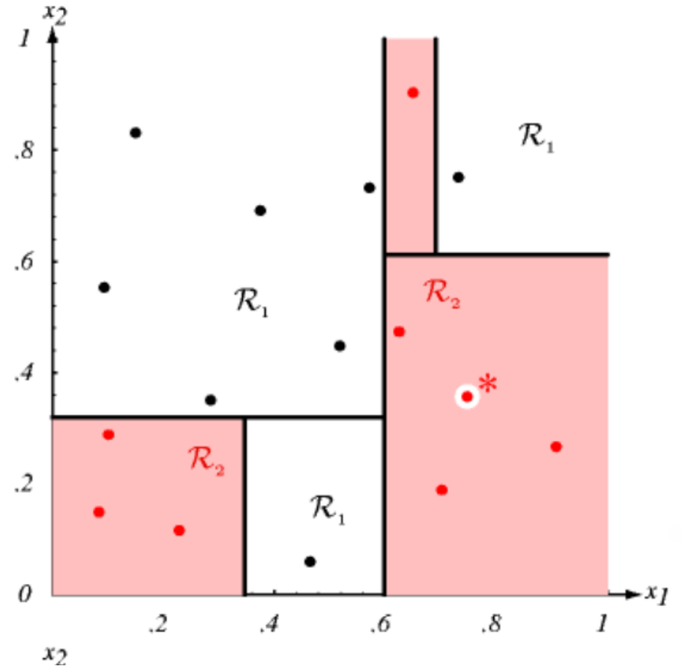

100
75
50
25
0
Drug Effectiveness (%)
0
10
20
30
40
Drug Dosage (mg)

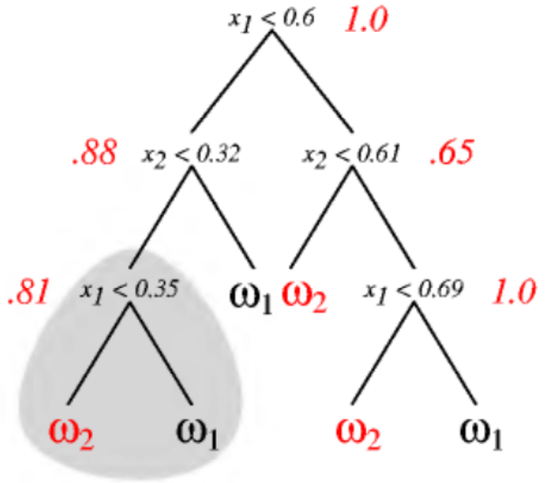

Dosage < 14.5
4.2% Effective
Dosage >= 29
2.5% Effective
Dosage >= 23.5
52.8% Effective
100% Effective

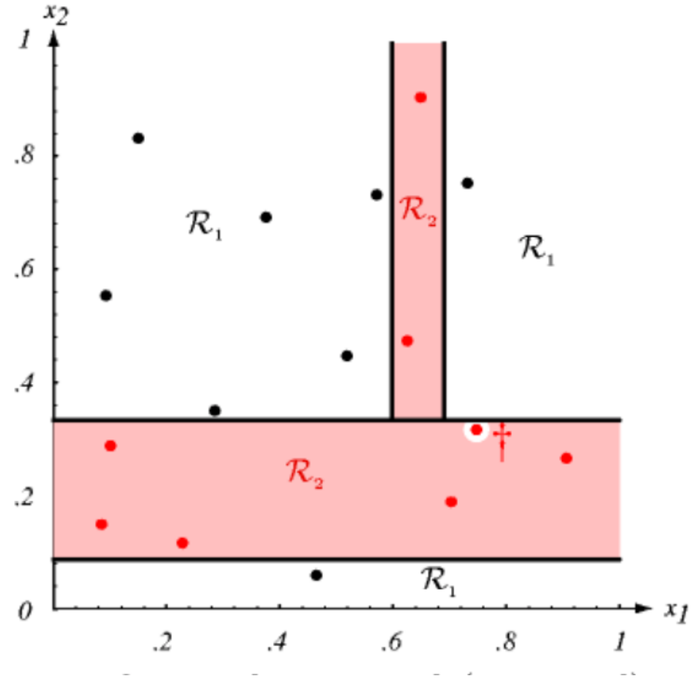
ω_1 (black)	
x_1	x_2
.15	.83
.09	.55
.29	.35
.38	.70
.52	.48
.57	.73
.73	.75
.47	.06

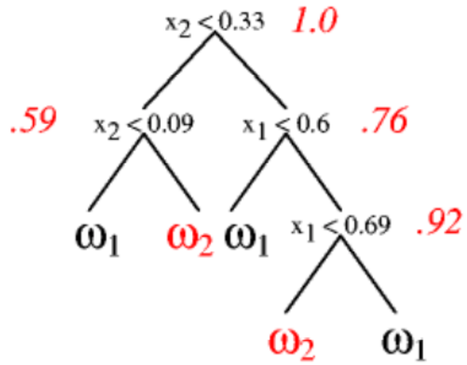
ω_2 (red)	
x_1	x_2
.10	.29
.08	.15
.23	.16
.70	.19
.62	.47
.91	.27
.65	.90
.75	.36* (.32†)

The instability is consequence of the discrete and greedy nature of tree creation









Random Forest

- Grow K different Trees
  - pick a random subset of training examples
  - pick d random attributes (compute gain based on subset instead of full set)
  - repeat K times