MDL
Assignment 4

2018111032
Vivek Pamnani

Problem/Dataset 3:

~~Horizontal Angle (degree)~~

Horizontal Angle (degree)	Distance (m)	Wind Speed (mph)	Kill
1.5	450	220	N
4.5	520	-120	Y
3	490	120	Y
5.5	530	117	Y
3.2	470	-170	N
5.2	505	-90	N
1.85	465	120	Y
4.8	517	147	Y
1.7	430	-100	Y

We have 9 objects in the dataset.
Info./bits req. for classification = $I_b(6, 3)$
$= 0.9183$

Attribute 'horizontalAngle':

Criterion: angle ≥ 1.6 (∵ gives 1 neat classification)

result: < 1.6 → 0p, 1n → I(0,1) = 0

≥ 1.6 → 6p, 2n → I(6,2) = 0.81128 (entropy = 0.7211)

By sorting,

1.6 seems optimal, raising the threshold ruins ~~at~~ both ~~neat sets~~ and resulting sets.

Gain = 0.9183 − 0.7211 = ~~0.107~~ **0.1972**

Attribute 'distance':

After sorting,

Criterion: distance ≥ 511 (∵ gives 1 neat classification)

result: < 511 → 3p, 3n → I(3,3) = 1

≥ 511 → 3p, 0n → I(3,0) = 0

gain: 0.9183 − $\left(\frac{6}{9} \times 1 + \frac{3}{9} \times 0\right)$ = ~~0.667~~ **0.2516**

(entropy)

Criterion: distance ≥ 480

result: ≥ 480 → 4p, 1n → I(4,1) = 0.722

< 480 → 2p, 2n → I(2,2) = 1

gain: 0.9183 − $\left(\frac{4}{9} \times 1 + \frac{5}{9} \times 0.722\right)$ = **0.073**

(entropy)

Attribute 'coined speed':

After sorting: speed

criterion: ≥ 103.5

result: ≥ 103.5 (4p, 1n) → I(4,1) = 0.722

< 103.5 (2p, 2n) → I(2,2) = 1

gain: 0.9183 − $\left(\frac{4}{9} \times 1 + \frac{5}{9} \times 0.722\right)$ = **0.0731**

Criterion: ≥ -95 (4p, 2n) → I(4,2) = 0.9183

< -95 (2p, 1n) → I(2,1) = 0.983

gain: 0.9183 − 0.9183 = **0**

∴ for first node, we pick the criterion:
'distance' ≥ 511 (gain: 0.2516)

Subset 1 (≥511)

4.5	520	-120	Y
5.5	530	117	Y
4.8	517	147	Y

⇒ Kill = Y

Subset 2 (<511)

I(3,3) = 1

Attribute 'horizontal angle':
After sorting,
Criterion: angle ≥ 3.1 0p, 2n → I(0,2) = 0
<3.1 3p, 1n → I(3,1) = 0.8113
gain: $1 - \left(\frac{4}{6} \times 0.8113 + \frac{2}{6} \times 0\right) = 0.4591$
(gives 1 neat classification)
other criterions degrade I(p,n) of both child sets (higher value) ⇒ lower gain

Attribut 'wind speed':
After sorting,
Criterion: speed ≥ 105 2p, 1n → I(2,1) = 0.9183
<105 1p, 2n → I(1,2) = 0.9183
gain: 1 − 0.9183 = 0.082

Criterion: speed ≥ -135 3p, 2n → I(3,2) = 0.97095
speed < -135 0p, 1n → I(0,1) = 0

gain: 1 − 0.80913 = 0.191

criteria: speed ≥ 170 → 0p, 1n → I(0,1) = 0
< 170 → 3p, 2n → I(3,2) = 0.971
gain: 1 − 0.8091 = 0.191

Attr: 'distance'
After sorting,
criterion: > 440
gain: 0.191
criterion: > 497.5
gain: 0.191

∴ for 2nd level, pick the criterion:
horizontal angle ≥ 3.1

subset 1 (≥ 3.1)

3.2	470	−170	N
5.2	505	−90	N

Kill = N

subset 2 (< 3.1)
I(3,1) = 0.8113

Attribute 'wind speed':
After sorting,
criterion: speed ≥ 170 → 0p, 1n → I(0,1) = 0
< 170 → 3p, 0n → I(3,0) = 0
gain = 0.8113 − 0 = 0.8113

higher gain not possible (∵ I ≠ 0)

For 3rd level, pick criterion:
~~Wind speed~~ ≥ 170

subset 1 (≥ 170)

Kill = N	1.5	450	220	N

subset 2 (< 170)

Kill = Y	1.7	430	-100	Y
	1.85	465	120	Y
	3	490	120	Y

Final Decision Tree: