⟨Decision Tree — Regression Trees⟩

- generalize decision tree to be regression algorithm ⇒ predict number not category

ex)

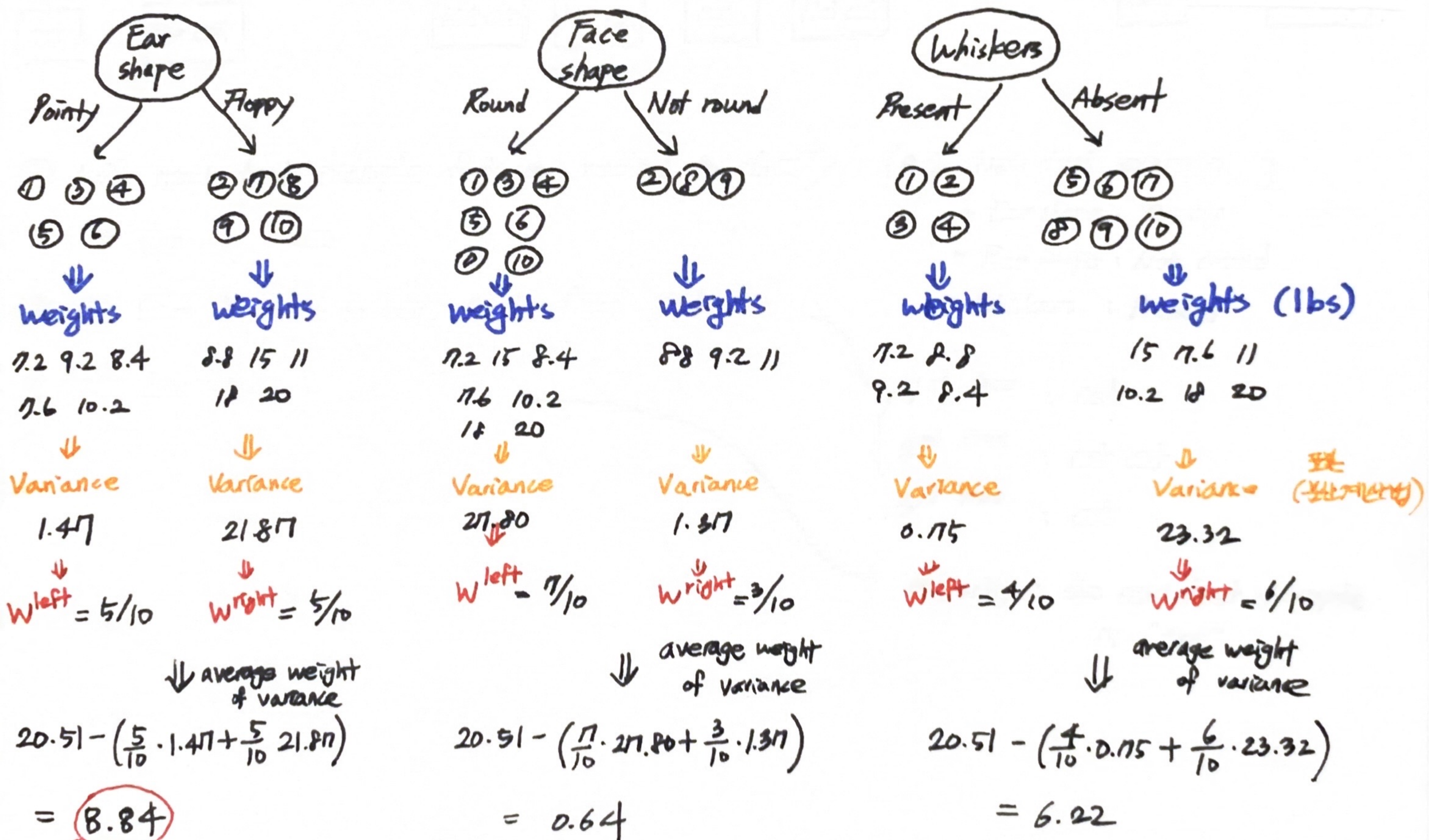|	Ear shape	Face shape	Whiskers	Weights (lbs)
①	Pointy	Round	Present	7.2
②	Floppy	Not round	Present	8.8
⋮	⋮	⋮	⋮	⋮
⑩	Floppy	Round	Absent	20

→ Variance (분산) = 20.51 (example)

$\underbrace{\qquad\qquad}_{x}$ $\underbrace{\qquad}_{y \text{ (target)}}$

* Choosing split ⇒ instead computing entropy, compute variance (분산)

Ear shape:
- Pointy: ① ③ ④ ⑤ ⑥ → weights: 7.2 9.2 8.4 7.6 10.2 → Variance 1.47 → $W^{left} = 5/10$
- Floppy: ② ⑦ ⑧ ⑨ ⑩ → weights: 8.8 15 11 18 20 → Variance 21.87 → $W^{right} = 5/10$

⇓ average weight of variance

$20.51 - \left(\frac{5}{10} \cdot 1.47 + \frac{5}{10} \cdot 21.87\right)$

= (8.84)

reduction in variance

↳ largest information gain

Face shape:
- Round: ① ③ ④ ⑤ ⑥ ⑦ ⑩ → weights: 7.2 15 8.4 7.6 10.2 18 20 → Variance 27.80 → $W^{left} = 7/10$
- Not round: ② ⑧ ⑨ → weights: 8.8 9.2 11 → Variance 1.37 → $W^{right} = 3/10$

⇓ average weight of variance

$20.51 - \left(\frac{7}{10} \cdot 27.80 + \frac{3}{10} \cdot 1.37\right)$

= 0.64

Whiskers:
- Present: ① ② ③ ④ → weights: 7.2 8.8 9.2 8.4 → Variance 0.75 → $W^{left} = 4/10$
- Absent: ⑤ ⑥ ⑦ ⑧ ⑨ ⑩ → weights: 15 7.6 11 10.2 18 20 → Variance 23.32 (분산이크다면) → $W^{right} = 6/10$

⇓ average weight of variance

$20.51 - \left(\frac{4}{10} \cdot 0.75 + \frac{6}{10} \cdot 23.32\right)$

= 6.22

→ with 5 examples each in left and right subsets of ear shape tree, build a new decision tree recursively (same as classification tree) until meeting the criteria for not splitting any further