Lecture 3.

- Recap
- Wrap up Decision Trees
- Overfitting ← come back Bias Variance Tradeoffs.
- Naive Bayes.

Recall :- For a set 'S', n_i = #examples w/ label 'i'.

$n_1 + \dots + n_\ell = n$ (ℓ = # possible labels)

$$H(S) := \underbrace{\sum_{i=1}^{\ell} \frac{n_i}{n} \log\left(\frac{n}{n_i}\right)}$$ (entropy of labels).

⇒ amt of randomness.

A - an attribute, takes values $\{v_1, \dots, v_a\}$.

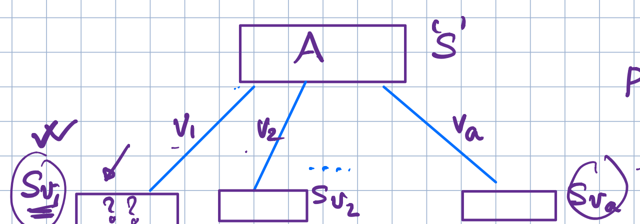
Information gain :- $IG(S; A)$

$$\boxed{IG(S;A) = H(S) - \sum_{j=1}^{a} \frac{|S_{v_j}|}{|S|} \cdot H(S_{v_j})}$$

S_{v_j}

initial randomness

randomness in examples where 'A' has value v_j



pure the better.

- Choose 'A' with highest $IG(S;A)$ ✓
- $\boxed{\text{RECURSIVE}}$ → until all examples pure keep building tree.

valent.

$$\xrightarrow{\text{minimize}} \sum_{j=1}^{a} \underbrace{\frac{|S_{v_j}|}{|S|}}_{\text{weight}} \cdot \underbrace{H(S_{v_j})}_{\text{(normalized)}}$$ average randomnes

Gini Impurity measure :- measure of impurity

For a set 'S', n_i = #examples w/label 'i'.

$n_1 + \dots + n_\ell = n$ (ℓ = # possible labels)

$$\text{Gini}(S) = 1 - \sum_{i=1}^{\ell} \left(\frac{n_i}{n}\right)^2 \qquad \boxed{n_i = \frac{n}{\ell}}$$

'n' $n_1, \dots, n_\ell$

If all examples in 'S' have same label, $\text{Gini}(S) = 0 \iff$ pure.

$$\text{Gini}(S;A) = \sum_{j=1}^{a} \frac{|S_{v_j}|}{n} \cdot \text{Gini}(S_{v_j}) \qquad \boxed{|S_{v_1}| + \dots + |S_{v_a}| = n}$$

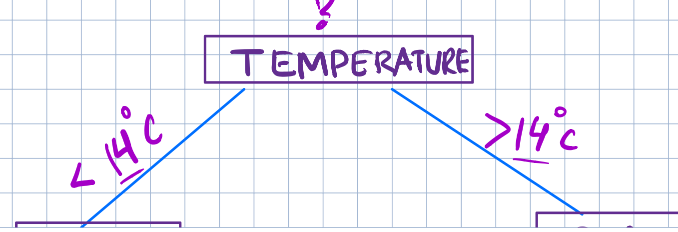
→ Pick 'A' with smallest $\text{Gini}(S;A)$.

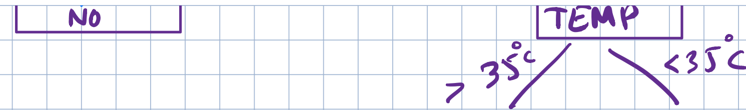
- Recursive defn.

Continuous attributes.

eg:- values of temperature insted of High/Normal.

Thresholding



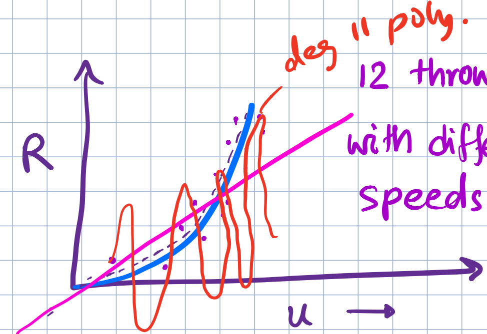


Overfitting.

- **OCCAM'S RAZOR** :- Simple models are good (small trees). why?

GOAL of learning :- Perform well on test data.

$$\boxed{R = \frac{u^2 \sin(2\theta)}{g}}$$



12 throws with different speeds.

Goal :- relate R, u, via a polynomial.

$$R = a_0 + a_1 u + a_2 u^2 + a_3 u^3 + \ldots$$

Ex

$$\boxed{n_1 + n_2 = 100}$$

Choose n_1, n_2 to min $\left(\frac{n_1}{n}\right)^2 + \left(\frac{n_2}{n}\right)^2 \longrightarrow$ minimize.

$n_1, n_2, n_3,$

GOAL :- Do well on unseen data.