

Decision Trees

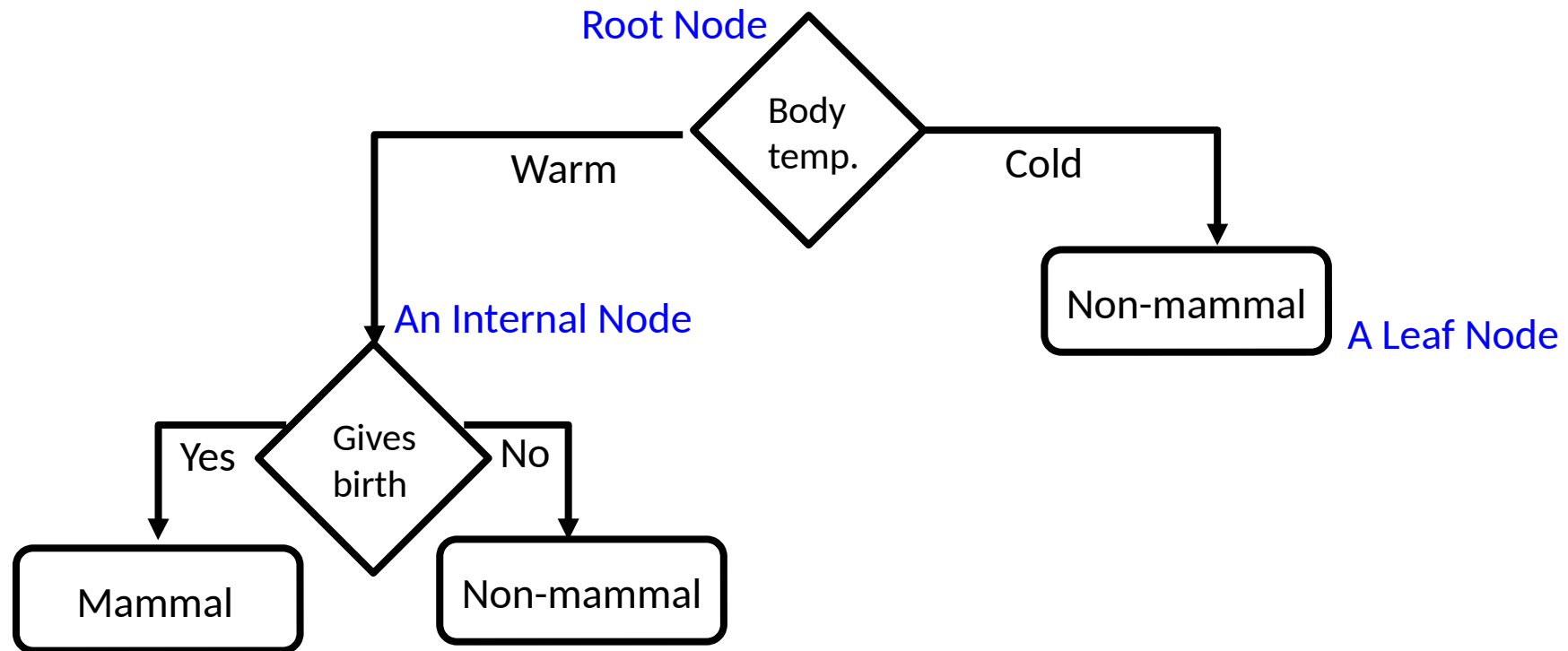


Decision Trees

- A Decision Tree (DT) defines a hierarchy of rules to make a prediction.
- Decision tree is an **nested if else condition classifier**.
- Root and internal nodes are Decision nodes. Leaf nodes make predictions.



Decision Trees



Decision Trees

1.Entropy:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log_b p(x),$$

- A measure for
 - uncertainty
 - purity
 - information content



Information gain

$$\text{Information Gain} = \text{Entropy (parent)} - \sum_{i=1}^n \frac{|S_i|}{|S|} \text{Entropy}(S_i)$$



Example dataset

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

