



Best Splits (Cont.)

- Measure of “cost” of a split:

$$\text{Information-content } (S, \{S_1, S_2, \dots, S_r\}) = - \sum_{i=1}^r \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

- **Information-gain ratio** =
$$\frac{\text{Information-gain } (S, \{S_1, S_2, \dots, S_r\})}{\text{Information-content } (S, \{S_1, S_2, \dots, S_r\})}$$
- The best split is the one that gives the maximum information gain ratio