How to pick nodes?

- A chosen attribute ***A***, with ***K*** distinct values, divides the training set ***E*** into subsets $E_1, \ldots, E_K$.
- The **Expected Entropy (EH) remaining** after trying attribute ***A*** (with branches $i=1,2,\ldots,K$) is

$$EH(A) = \sum_{i=1}^{K} \frac{p_i + n_i}{p + n} H\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

- **Information gain (I)** or **reduction in entropy** for this attribute is:

$$I(A) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - EH(A)$$

= Entropy in the **parent node -** remaining **Expected Entropy** in the **child nodes**

[Hwee Tou Ng & Stuart Russell]

Split = which node to split ⇒ stop? Yes / No
by what attribute? (feat)
→ by attribute that gives best ΔE (IG)
GREEDY
not optimal for best tree overall

points in child i

at a given node (data)
- Decide = split
- Try BRUTE FORCE every attribute, see which is best