



---

**算法 1** 信息增益算法
 

---

**输入:** 训练数据集 $D$ 和特征 $A$

**输出:** 特征 $A$ 对训练数据集 $D$ 的信息增益 $g(D, A)$

(1) 计算数据集 $D$ 的经验熵 $H(D)$

$$H(D) = - \sum_{k=1}^K \frac{|C_k|}{|D|} \log_2 \frac{|C_k|}{|D|}$$

(2) 计算特征 $A$ 对数据集 $D$ 的经验条件熵 $H(D|A)$

$$H(D|A) = \sum_{i=1}^n \frac{|D_i|}{|D|} H(D_i) = - \sum_{i=1}^n \frac{|D_i|}{|D|} \sum_{k=1}^K \frac{|D_{ik}|}{|D_i|} \log_2 \frac{|D_{ik}|}{|D_i|}$$

(3) 计算信息增益

$$g(D|A) = H(D) - H(D|A)$$


---