۹۹۱۳۰۰۴ بهراد اعتضادی

$$* \; H(X) = E\left(-\log_2(P(X))\right) = -\sum_{i=1}^{n} P(x_i)\log_2(P(x_i)).$$

$$* \; Gain(S) = H(X) - \sum_{i=1}^{n} \frac{|X_i|}{|X|} H(X_i)$$

* تعریف می‌کنیم: مجموعه داده برچسب خورده X و ویژگی که براساس آن جداسازی می‌کنیم K باشد

آنگاه $P(X,K)$ تابع چگالی توأم X و K است

پس داریم: $P(X) = \sum_K P(X,K)$, $P(K) = \sum_X P(X,K)$

$$\rightarrow Gain(X,K) = H(X) - H(X|K)$$

$$\rightarrow Gain(X,K) = -\sum_{X} P(X)\log_2(P(X)) - \sum_{K} P(K)\sum_{X}\left(-P(X|K)\log_2(P(X|K))\right)$$

$$\rightarrow -Gain(X,K) = \sum_{X}\sum_{K} P(X,K)\log_2 P(X) - \sum_{K}\sum_{X}\left(P(X,K)\log_2(P(X|K))\right)$$

$$\rightarrow -Gain(X,K) = \sum_{X}\sum_{K} P(X,K)\left(\log_2(P(X)) - \log_2(P(X|K))\right)$$

$$\rightarrow -Gain(X,K) = \sum_{X}\sum_{K} P(X,K)\left(\log_2\left(\frac{P(X)}{P(X|K)}\right)\right)$$

$$\rightarrow -Gain(X,K) = \sum_{X}\sum_{K} P(X|K)P(K)\left(\log_2\left(\frac{P(X)}{P(X|K)}\right)\right)$$

$$\rightarrow -Gain(X,K) = \sum_{K} P(K)\sum_{X} P(X|K)\left(\log_2\left(\frac{P(X)}{P(X|K)}\right)\right)$$

$$\sum_{i=1}^{v} a_i \log x_i \leqslant \log\left(\sum_{i=1}^{v} a_i x_i\right)$$: Jensen نامساوی

$$a_i \geqslant 0 \,, \quad \sum_i a_i = 1$$

$$\longrightarrow -\mathrm{Gain}(X,k) \leqslant \sum_k P(k)\left(\log_2\left(\sum_X \frac{P(X|k)P(X)}{P(X|k)}\right)\right)$$

$$\longrightarrow -\mathrm{Gain}(X,k) \leqslant \log_2\left(\sum_k \sum_X \left(\frac{P(k)P(X|k)P(X)}{P(X|k)}\right)\right)$$

$$\longrightarrow -\mathrm{Gain}(X,k) \leqslant \log_2\left(\sum_k \sum_X P(k)P(X)\right)$$

$$\longrightarrow -\mathrm{Gain}(X,k) \leqslant \log_2\left(\sum_k P(k) \sum_X P(X)\right)$$

$$\longrightarrow -\mathrm{Gain}(X,k) \leqslant \log_2\left(\sum_k P(k)\right)$$

$$\longrightarrow -\mathrm{Gain}(X,k) \leqslant \log_2(1) = 0$$

$$\longrightarrow \underline{\mathrm{Gain}(X,k) \geqslant 0}$$

پس ثابت می‌شود بهره همواره نامنفی است.