

Calculo de ganancia de la información

La cantidad de información, medida en bits, producida por la ocurrencia de un evento es inversa a la probabilidad de la ocurrencia de dicho evento.

$$I(P(V_1) + \dots + P(V_n)) = -\sum_{i=1}^n P(V_i) \cdot \log_2 P(V_i)$$

Promedio de información de los diferentes eventos

Es decir que la ganancia de un evento A cualquiera puede ser expresado como:

$$G(A) = I - E(A)$$

La entropía del evento A

La cantidad total de información

$$G(A) = P\left(\frac{p}{p+n}; \frac{n}{p+n}\right) - E(A)$$

$$\text{Donde: } P\left(\frac{p}{p+n}; \frac{n}{p+n}\right) = -\frac{p}{p+n} \cdot \log_2 \frac{p}{p+n} - \frac{n}{p+n} \cdot \log_2 \frac{n}{p+n}$$

$$E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} \cdot I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

Ejemplo utilizando la tabla de león o no león

Peludo	Edad	Tamaño	Clase
SI	VIEJO	GRANDE	LEON
NO	JOVEN	GRANDE	NO LEON
SI	JOVEN	MEDIANO	LEON
SI	VIEJO	PEQUEÑO	NO LEON
SI	JOVEN	PEQUENO	NO LEON
SI	JOVEN	GRANDE	LEON
NO	JOVEN	PEQUENO	NO LEON
NO	VIEJO	GRANDE	NO LEON

Calculando la información total:

$$I(p;n) = -\frac{p}{p+n} \cdot \log_2 \frac{p}{p+n} - \frac{n}{p+n} \cdot \log_2 \frac{n}{p+n}$$

$$I(p;n) = -\frac{3}{8} \cdot \log_2 \frac{3}{8} - \frac{5}{8} \cdot \log_2 \frac{5}{8}$$

$$I(p;n) = 0,531 + 0,884 = 0,9544$$

Calculando la información y entropía para el atributo peludo:

Evento SI: $\frac{p_1=3}{n_1=2} \rightarrow I(p_1;n_1) \equiv I(3;2) = -\frac{3}{5} \cdot \log_2 \frac{3}{5} - \frac{2}{5} \cdot \log_2 \frac{2}{5} = 0,971$

Evento NO: $\frac{p_2=0}{n_2=3} \rightarrow I(p_2;n_2) \equiv I(0;3) = -\frac{0}{3} \cdot \log_2 \frac{0}{3} - \frac{3}{3} \cdot \log_2 \frac{3}{3} = 0$

Entropía de peludo:

$$E(\text{peludo}) = \sum_{i=1}^v \frac{p_i+n_i}{p+n} \cdot I(p_i;n_i)$$

$$E(\text{peludo}) = \frac{5}{8} \cdot 0,971 + \frac{3}{8} \cdot 0 = 0,607$$

Ganancia de peludo:

$$G(\text{peludo}) = I - E(\text{peludo}) = 0,9544 - 0,607$$

$$G(\text{peludo}) = 0,3475$$

Calculando la información y entropía para el atributo edad:

Evento VIEJO: $\frac{p_1=1}{n_1=2} \rightarrow I(p_1;n_1) \equiv I(1;2) = -\frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{2}{3} \cdot \log_2 \frac{2}{3} = 0,918$

Evento JOVEN: $\frac{p_2=2}{n_2=3} \rightarrow I(p_2;n_2) \equiv I(2;3) = -\frac{2}{5} \cdot \log_2 \frac{2}{5} - \frac{3}{5} \cdot \log_2 \frac{3}{5} = 0,971$

Entropía de edad:

$$E(\text{edad}) = \sum_{i=1}^v \frac{p_i+n_i}{p+n} \cdot I(p_i;n_i)$$

$$E(\text{edad}) = \frac{3}{8} \cdot 0,918 + \frac{5}{8} \cdot 0,971 = 0,951$$

Ganancia de edad:

$$G(\text{edad}) = I - E(\text{edad}) = 0,9544 - 0,951$$

$$G(\text{edad}) = 0,0033$$

Calculando la información y entropía para el atributo tamaño:

Evento GRANDE: $p_1=2 \rightarrow I(p_1; n_1) \equiv I(2; 2) = -\frac{2}{4} \cdot \log_2 \frac{2}{4} - \frac{2}{4} \cdot \log_2 \frac{2}{4} = 1$
 $n_1=2$

Evento MEDIANO: $p_2=1 \rightarrow I(p_2; n_2) \equiv I(1; 0) = -\frac{1}{1} \cdot \log_2 \frac{1}{1} - \frac{0}{1} \cdot \log_2 \frac{0}{1} = 0$
 $n_2=0$

Evento PEQUEÑO: $p_3=0 \rightarrow I(p_3; n_3) \equiv I(0; 3) = -\frac{0}{3} \cdot \log_2 \frac{0}{3} - \frac{3}{3} \cdot \log_2 \frac{3}{3} = 0$
 $n_3=3$

Entropía de tamaño:

$$E(\text{tamaño}) = \sum_{i=1}^v \frac{p_i + n_i}{p+n} \cdot I(p_i; n_i)$$
$$E(\text{tamaño}) = \frac{4}{8} \cdot 1 + \frac{1}{8} \cdot 0 + \frac{3}{8} \cdot 0 = 0,5$$

Ganancia de tamaño:
$$\boxed{G(\text{tamaño}) = I - E(\text{tamaño}) = 0,9544 - 0,5}$$
$$G(\text{tamaño}) = 0,4544$$

Escala de Ganancia:

- 1. G(tamaño)=0,4544**
- 2. G(peludo)=0,3475**
- 3. G(edad)=0,0033**