

Treesformulas.pdf



Beaflm



Inteligencia Artificial



3º Grado en Ingeniería Informática



**Escuela Politécnica Superior
Universidad Autónoma de Madrid**

BBVA**1/6**

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

Ventajas Cuenta Online de BBVA

0€

Sin comisión de administración o mantenimiento de cuenta.
(0 % TIN 0 % TAE)

0€

Sin comisión por emisión y mantenimiento de Tarjeta Aqua débito.

0

Sin necesidad de domiciliar nómina o recibos.

Las ventajas de **WUOLAH PRO**



Di adiós a la publi en los apuntes y en la web



Descarga carpetas completas de un tirón



Acumula tickets para los sorteos

cómo??





1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

ventajas

PRO



Di adiós a la publi en los apuntes y en la web



Acumula tickets para los sorteos



Descarga carpetas completas

estudia sin publi
WUOLAH PRO

TREES

- Gini impurity

$$\text{Gini}(A) = 1 - \sum_{i=1}^M P(A_i)^2 = 1 - \sum_{i=1}^M \left(\frac{\text{Freq}(A_i)}{N} \right)^2$$

→ Used in binary

→ simplification : $2 * P(\text{class1}) * P(\text{class2})$

- ID3 :

- Entropy : lack of homogeneity of a set of examples with respect to their class

- Information Gain (IG): difference between the entropy of the original set and all the subsets obtained.

$$\text{IG} = \text{Entropy decrease}_A(x) = E(x) - E_A(x)$$

↳ In each node, the attribute that causes the greatest decrease is chosen.

$$H(A) = - \sum_{i=1}^M P(A_i) \log P(A_i) = - \sum_{i=1}^M \frac{\text{Freq}(A_i)}{N} \log \frac{\text{Freq}(A_i)}{N}$$

Conditional Entropy of C/A =

$$H(C/A) = \sum_{i=1}^M P(A_i) \times H(C/A_i) = \\ = \sum_{i=1}^M \frac{\text{Freq}(A_i)}{N} \times H(C/A_i)$$

if $A = A_i$:

$$H(C/A_i) = - \sum_{j=1}^L P(C_j/A_i) \log_2 P(C_j/A_i) = \\ = - \sum_{j=1}^L \frac{\text{Freq}(C_j)}{K} \log_2 \frac{\text{Freq}(C_j)}{K}$$

↳ número de veces que A_i aparece en los datos

$$IG(C/A) = H(C) - H(C/A)$$

$$\text{↳ ex: } IG(\text{Decision} / \text{wind}) = \\ H(\text{Decision}) - H(\text{Decision} / \text{Wind})$$

• C4.5 : evolution of ID3

- Handles also continuous values

- Uses IG and IG Ratio

Ábrete la Cuenta Online de BBVA y llévate 1 año de Wuolah PRO

Cómo??



Las ventajas de **WUOLAH PRO**



Di adiós a la publi en los apuntes y en la web



Descarga carpetas completas de un tirón



Acumula tickets para los sorteos

Ventajas Cuenta Online de BBVA

0€

Sin comisión de administración o mantenimiento de **cuenta**.
(0 % TIN 0 % TAE)

0€

Sin comisión por emisión y mantenimiento de **Tarjeta** Aqua débito.

0

Sin necesidad de domiciliar nómina o recibos.

$$IGR(C/A) = \frac{IG(C/A)}{IV(A)}$$

$IV(A) = H(A)$ \rightsquigarrow La entropía de la clase!

- How to handle a continuous value?

Binarize the series!

1. Order the values $\{3, 1, 2\} \rightarrow \{1, 2, 3\}$

2. Choose thresholds: $\{1, 2, 3\} \rightarrow \begin{matrix} x > 1, x \leq 1 \\ x > 2, x \leq 2 \end{matrix}$

3. Treat them as binary:

- Count them as separate groups
- Calculate entropy for each of them

4. Calculate the IG of the thresholds and choose the best

5. If IGR, now compute $IV(\text{Class})$
 $IV(\text{Age} \neq 2)$, for example

• Trees: Adv vs Drawbacks

- ✓ Simple Implementation
- ✓ Very little data preparation
- ✓ Variable selection
- ✓ Interpretable results
- ✓ Fast training & prediction



1/6

Este número es indicativo del riesgo del producto, siendo 1/6 indicativo de menor riesgo y 6/6 de mayor riesgo.

BBVA está adherido al Fondo de Garantía de Depósitos de Entidades de Crédito de España. La cantidad máxima garantizada es de 100.000 euros por la totalidad de los depósitos constituidos en BBVA por persona.

ventajas

PRO



Di adiós a la publi en los apuntes y en la web



Acumula tickets para los sorteos



Descarga carpetas completas

estudia sin publi

WUOLAH PRO

✓ Non-linear relationships do not affect per y.

- X Not very accurate predictions
- X Low variance → rigid models
- X However, they can be used as bases for an ensemble

• KNN

- We encode the variables
- For example, if a class has only T & F,
 $T=1$ & $F=0$.
- If a column has 3 options, we can divide them and turn them into T/F.

