

# Árboles de Decisión

Cristian López Del Alamo

[clopezd@unsa.edu.pe](mailto:clopezd@unsa.edu.pe)

IPRODAM3D - Research group

Mayo, 2022



# 1

## Árboles de Decisión



**UNSA**  
UNIVERSIDAD NACIONAL DE SAN AGUSTÍN DE AREQUIPA



## Base de Datos: Vampiros

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
Si	Si	Rojo	No	No
?	No	Rojo	No	Si
No	No	Promedio	Fuerte	Si
?	No	Promedio	Extraño	Si
Si	No	Pálido	Fuerte	No
Si	No	Promedio	Fuerte	No
?	Si	Rojo	Extraño	No

## Base de Datos: Vampiro

F1

F2

F3

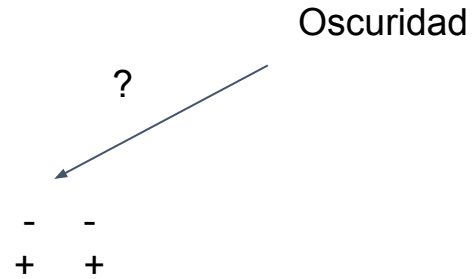
F4

Y

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
Si	Si	Rojo	No	No
?	No	Rojo	No	Si
No	No	Promedio	Fuerte	Si
?	No	Promedio	Extraño	Si
Si	No	Pálido	Fuerte	No
Si	No	Promedio	Fuerte	No
?	Si	Rojo	Extraño	No

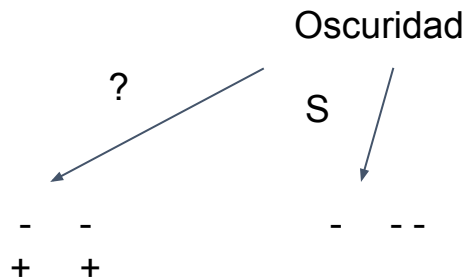
## Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



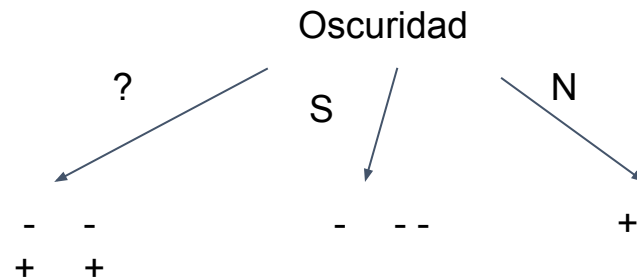
## Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



## Analizamos F1 : Oscuridad

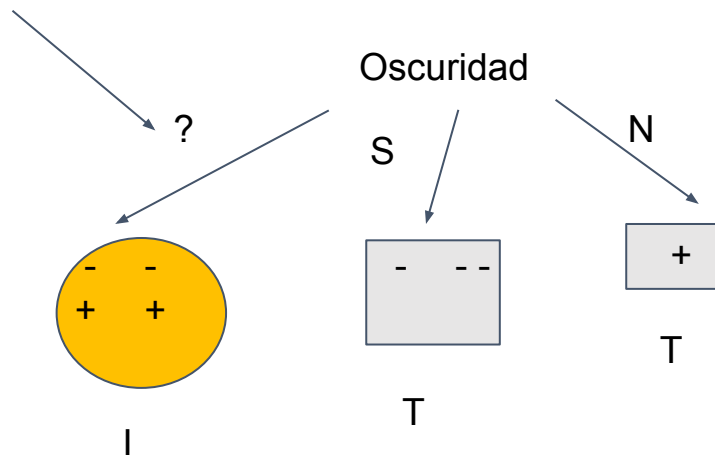
Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No





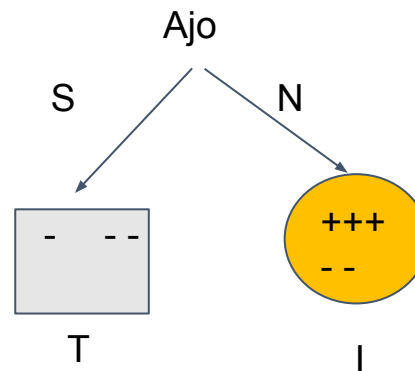
## Analizamos F1 : Oscuridad

Oscuridad	Vampiro
?	No
Si	No
?	Si
No	Si
?	Si
Si	No
Si	No
?	No



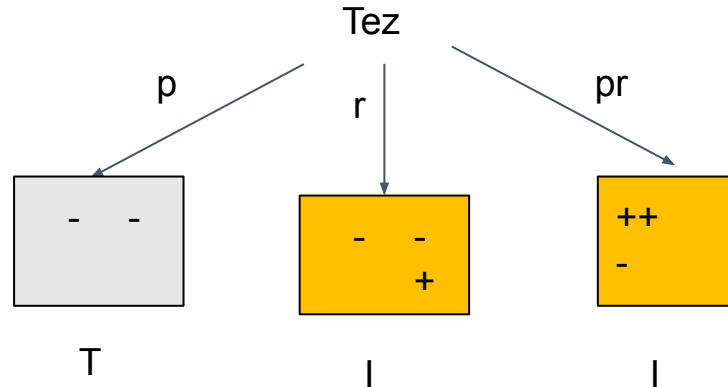
## Analizamos F2 : Ajo

F2	Y
Ajo	Vampiro
Si	No
Si	No
No	Si
No	Si
No	Si
No	No
No	No
Si	No



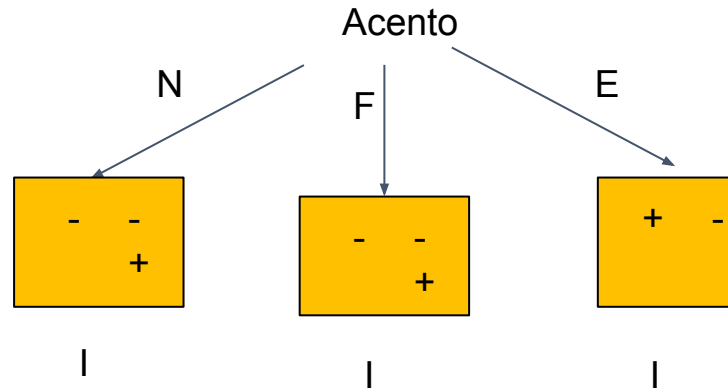
## Analizamos F3 : Tez

F3	Y
Tez	Vampiro
Pálido	No
Rojo	No
Rojo	Si
Promedio	Si
Promedio	Si
Pálido	No
Promedio	No
Rojo	No

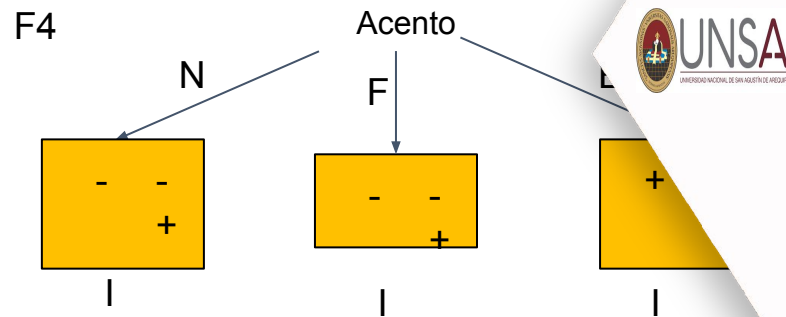
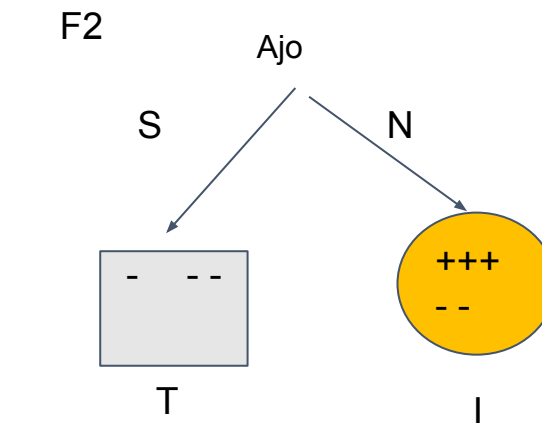
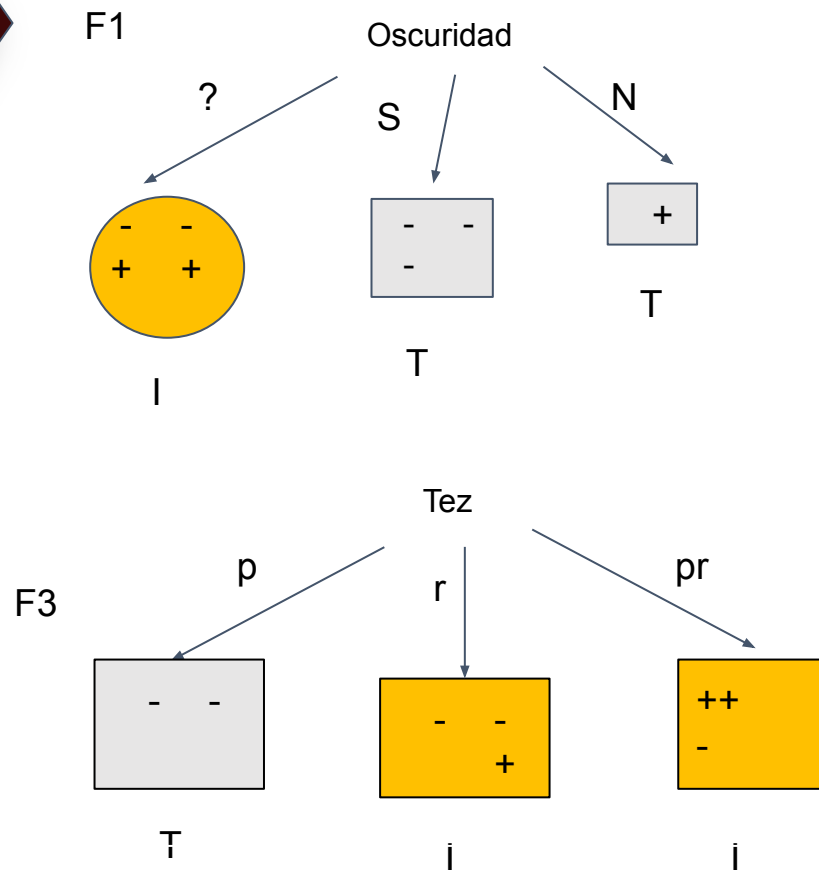


## Analizamos F4 : Ajo

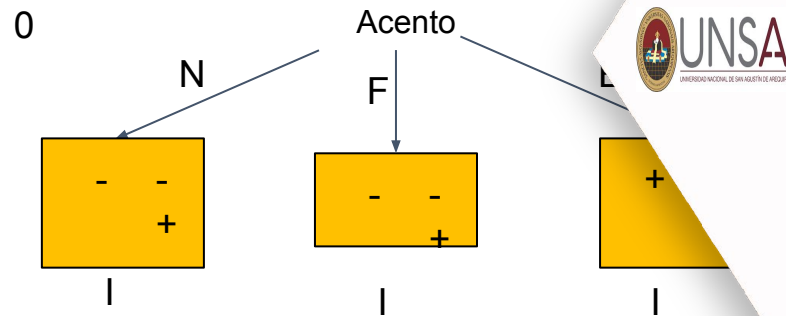
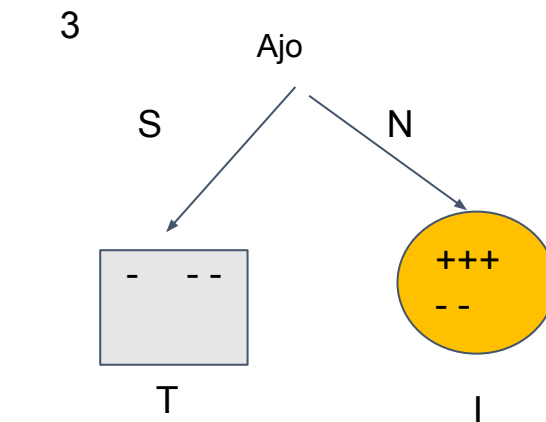
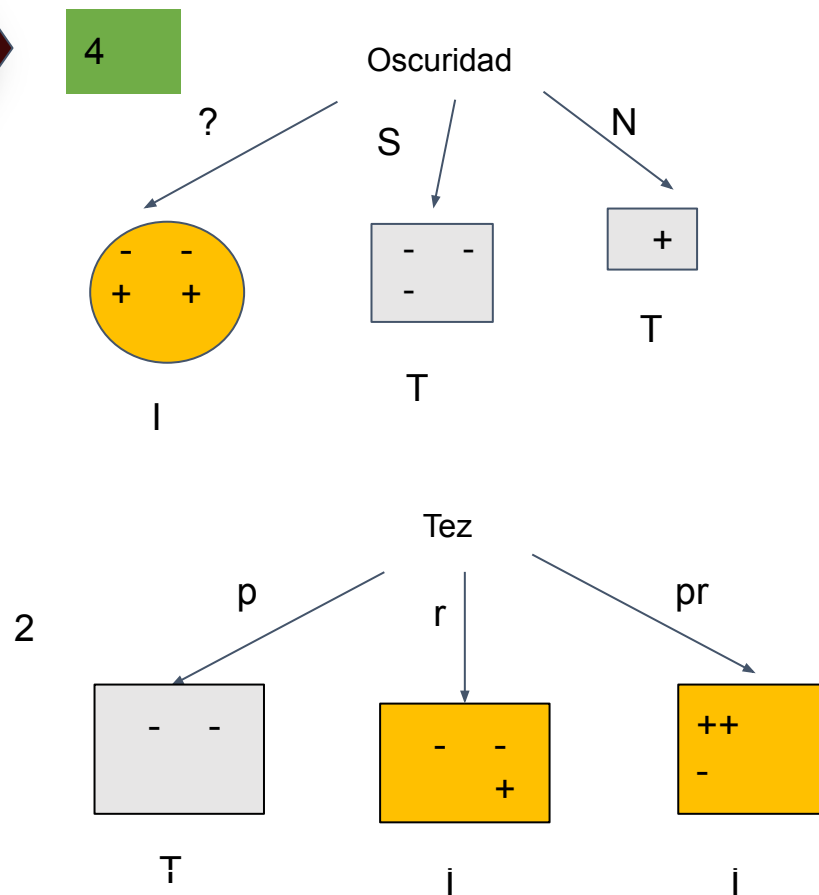
F4	Y
Acento	Vampiro
No	No
No	No
No	Si
Fuerte	Si
Extraño	Si
Fuerte	No
Fuerte	No
Extraño	No



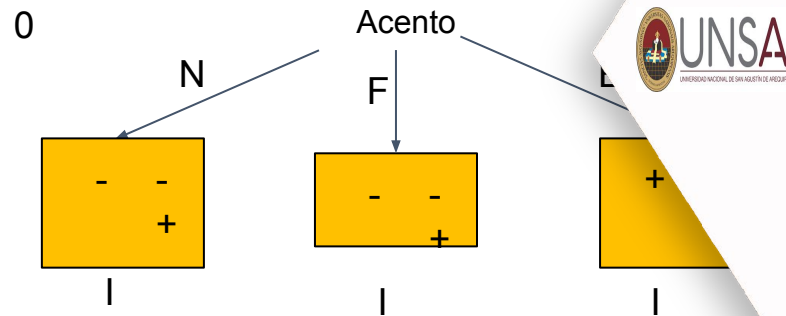
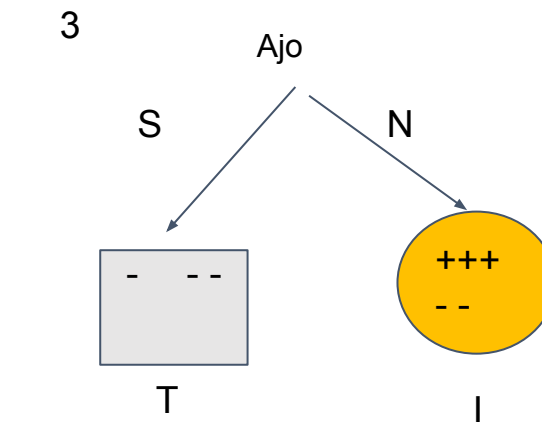
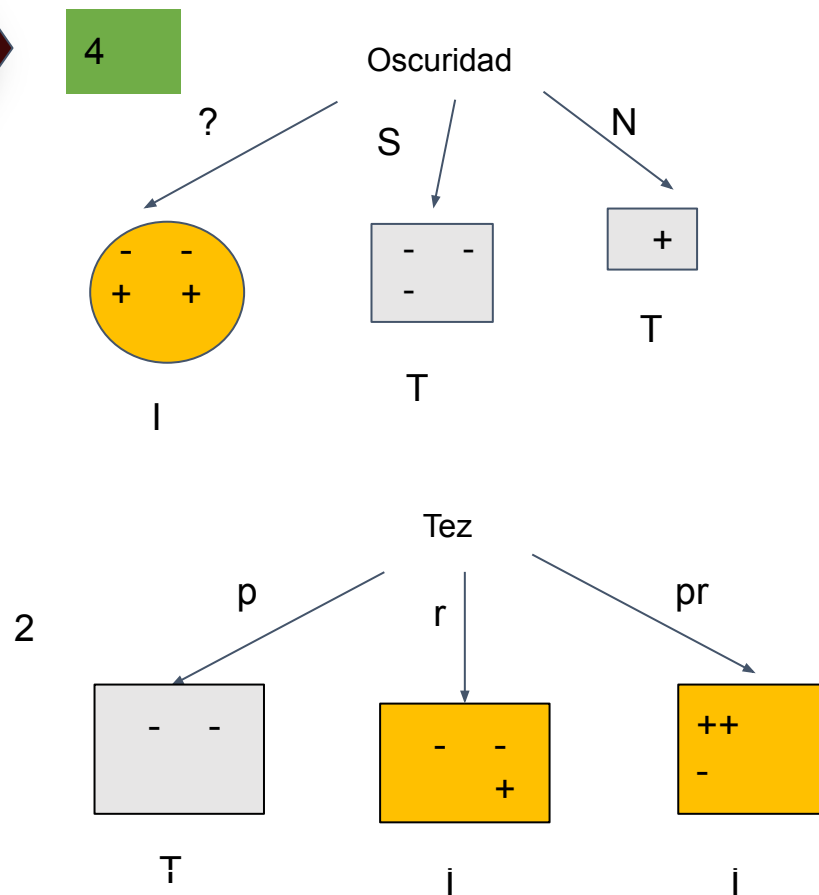
## Analizando las 4 características



# Analizando las 4 características



# Analizando las 4 características



Analizamos los nodos potenciales, internos o prometedores

Ajo	Tez	Acento	Vampiro
Si	Pálido	No	No
Si	Rojo	No	No
No	Rojo	No	Si
No	Promedio	Fuerte	Si
No	Promedio	Extraño	Si
No	Pálido	Fuerte	No
No	Promedio	Fuerte	No
Si	Rojo	Extraño	No

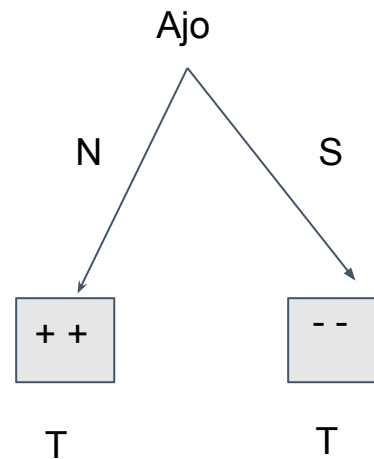


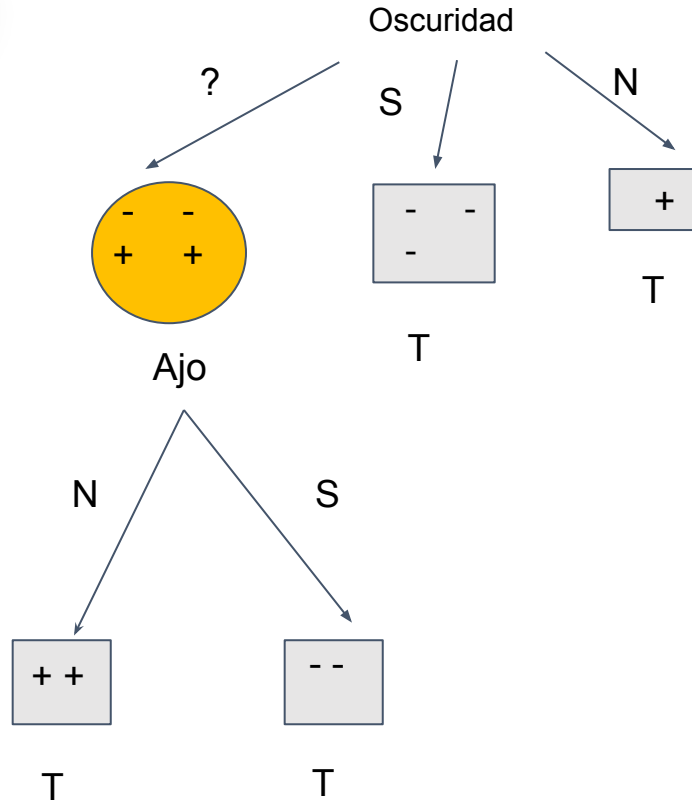
F2

Oscuridad	Ajo	Tez	Acento	Vampiro
?	Si	Pálido	No	No
?	No	Rojo	No	Si
?	No	Promedio	Extraño	Si
?	Si	Rojo	Extraño	No

Oscuridad	Ajo	Vampiro
?	Si	No
?	No	Si
?	No	Si
?	Si	No

Oscuridad	Ajo	Vampiro
?	Si	No
?	No	Si
?	No	Si
?	Si	No





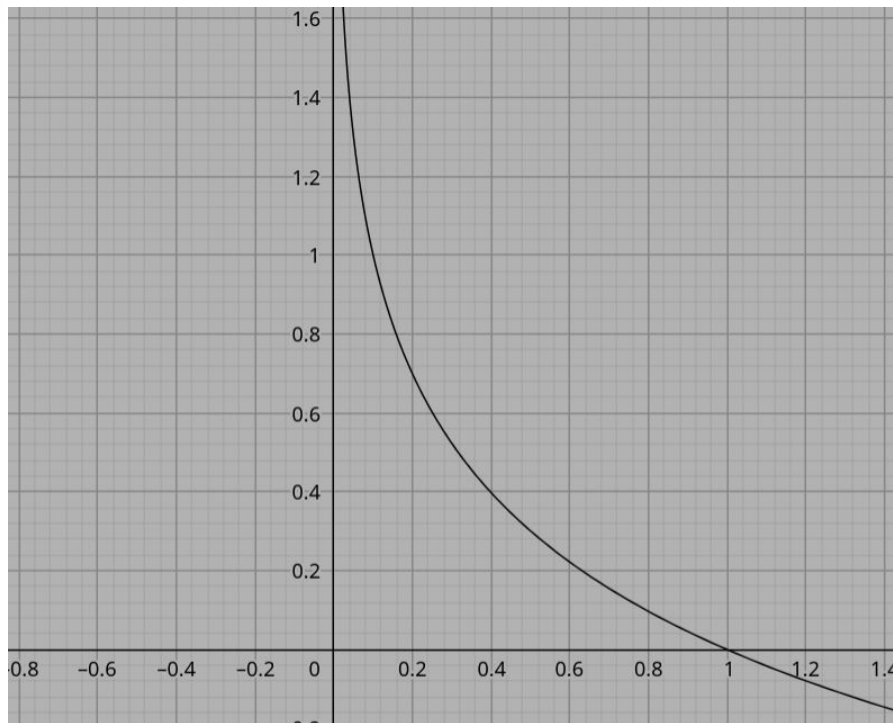
Si Oscuridad == S  
     return No Vampiro  
 Si Oscuridad == N  
     return Vampiro  
 Si Oscuridad == ?  
     Si Ajo == N  
         return Vampiro  
     Sino  
         return No Vampiro

¿Qué Problemas encuentra con este método?

# 2

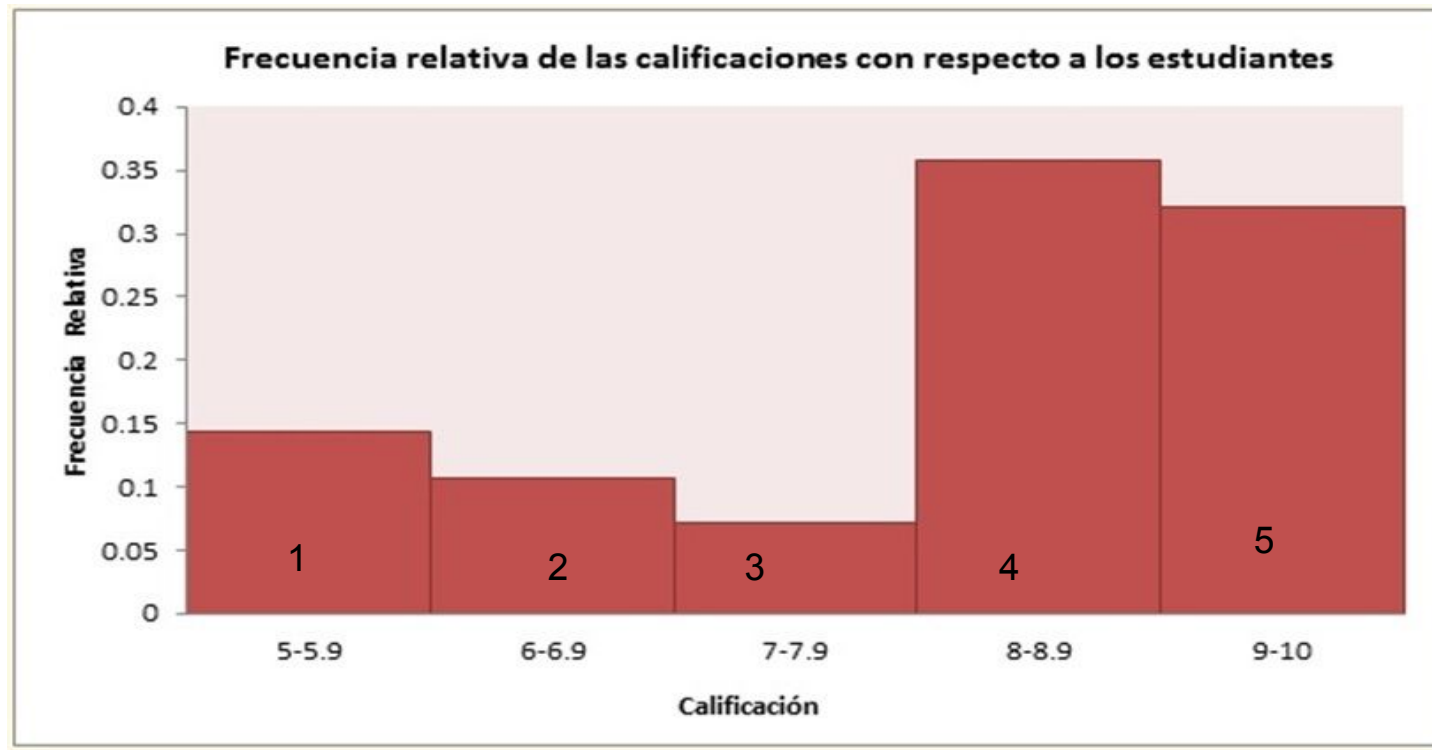
## Información y Entropía

# Información



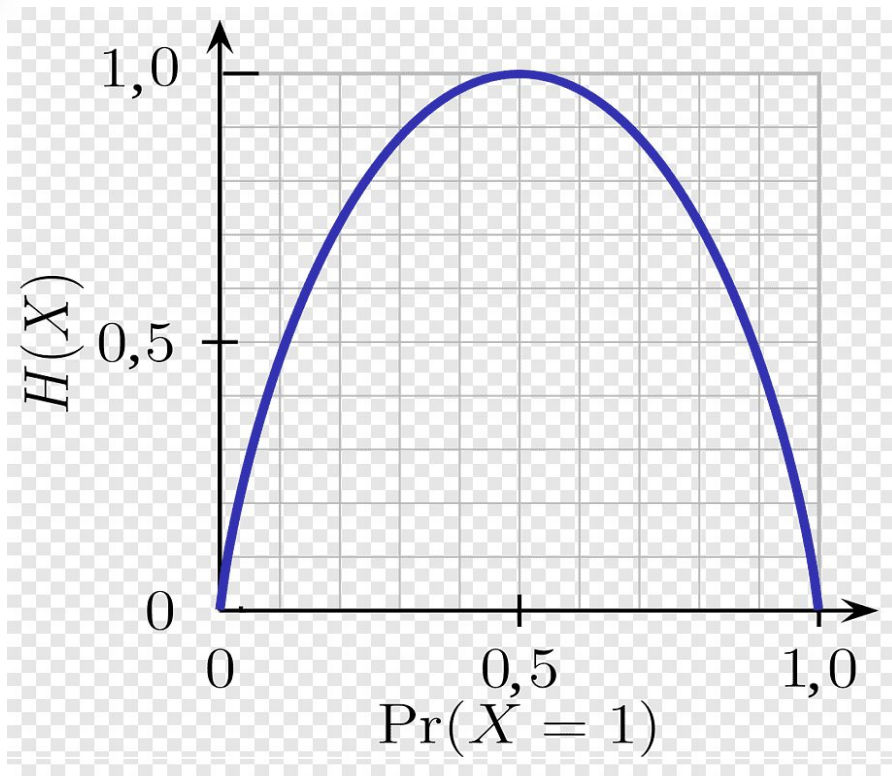
$$I(X) = \log(1/p(x)) = -\log(p(x))$$

¿Qué grupo tiene más información?





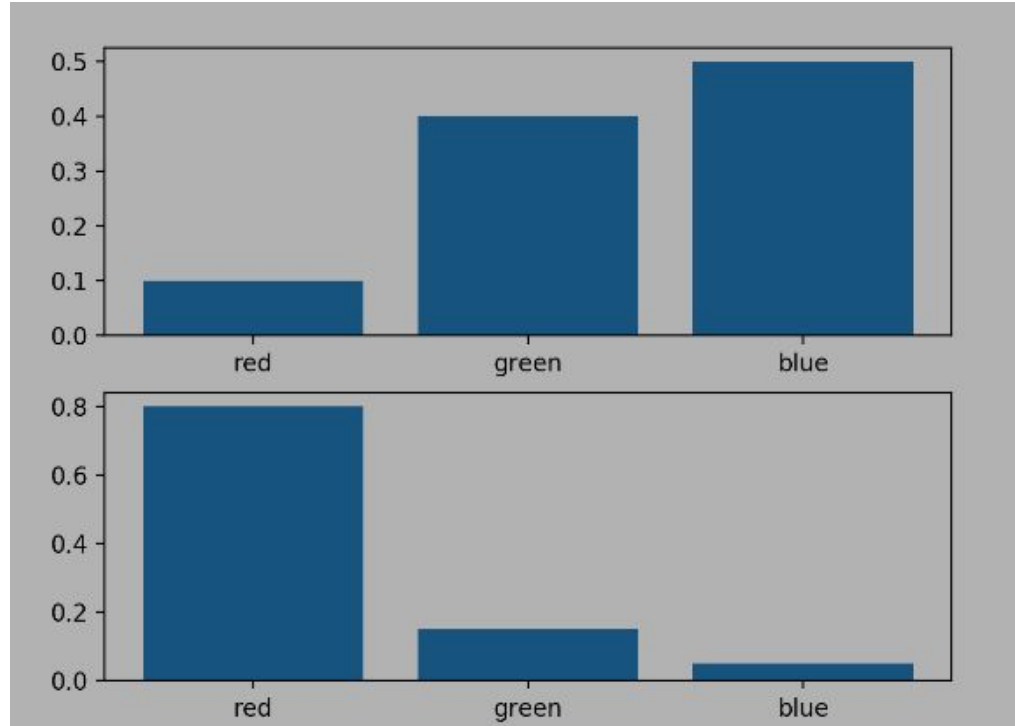
# Entropía



$$H(X) = E[I(X)] = \sum_{i=1}^n p(x_i) I(x_i) = - \sum_{i=1}^n p(x_i) \log(x_i)$$

## ¿ Cual de los dos histogramas tiene mayor entropía?

1

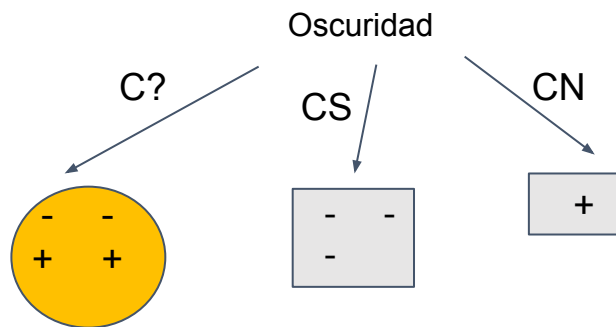


2

Nivel de Desorden de un conjunto o una clase

$$D_s = - \sum_{c \in C} P_c \log_2 P_c$$

¿Cuál es el nivel de desorden de las 3 clases en Oscuridad?

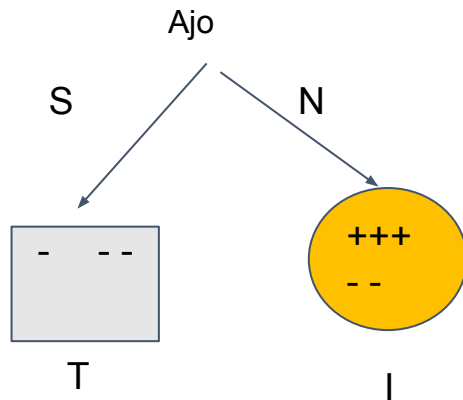


$$D(C?) =$$

$$D(CS) =$$

$$D(CN) =$$

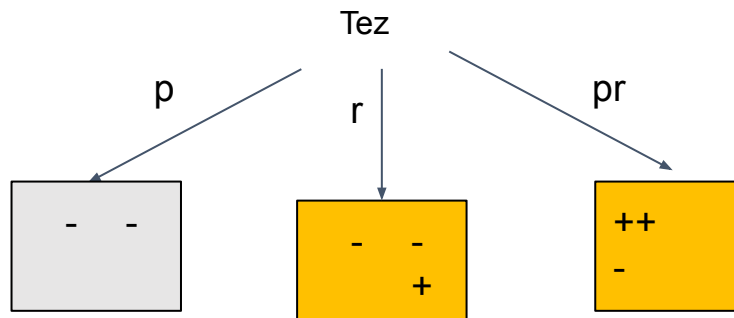
¿Cuál es el nivel de desorden de las 3 clases en Ajo?



$D(CS) =$

$D(CN) =$

¿Cuál es el nivel de desorden de las 3 clases en Ajo?

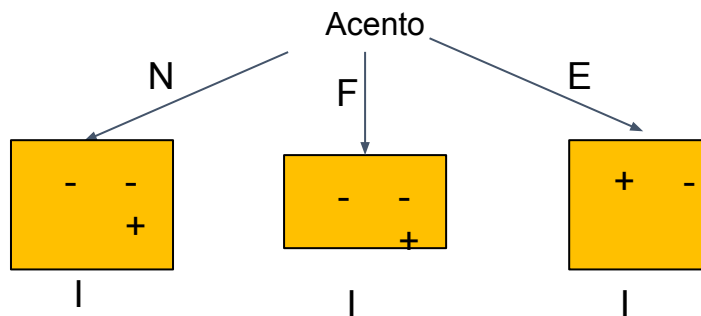


$D(C_p) =$

$D(C_r) =$

$D(C_{pr}) =$

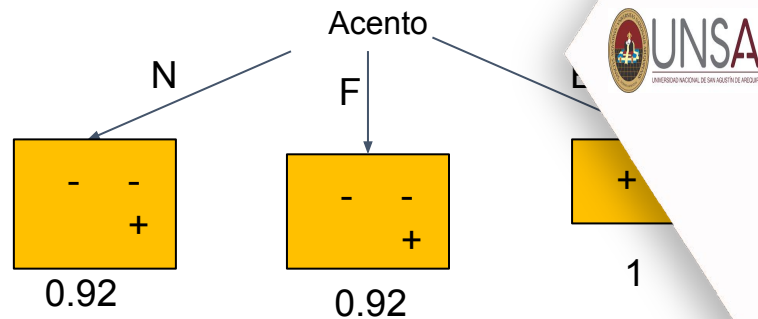
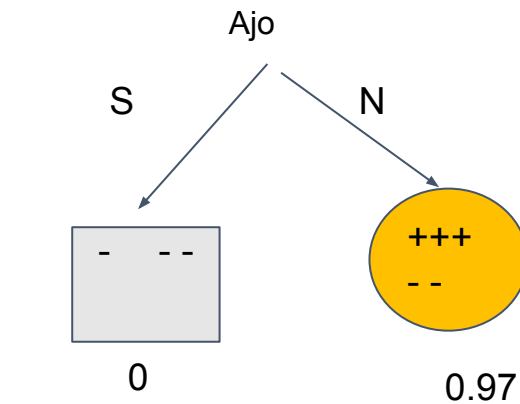
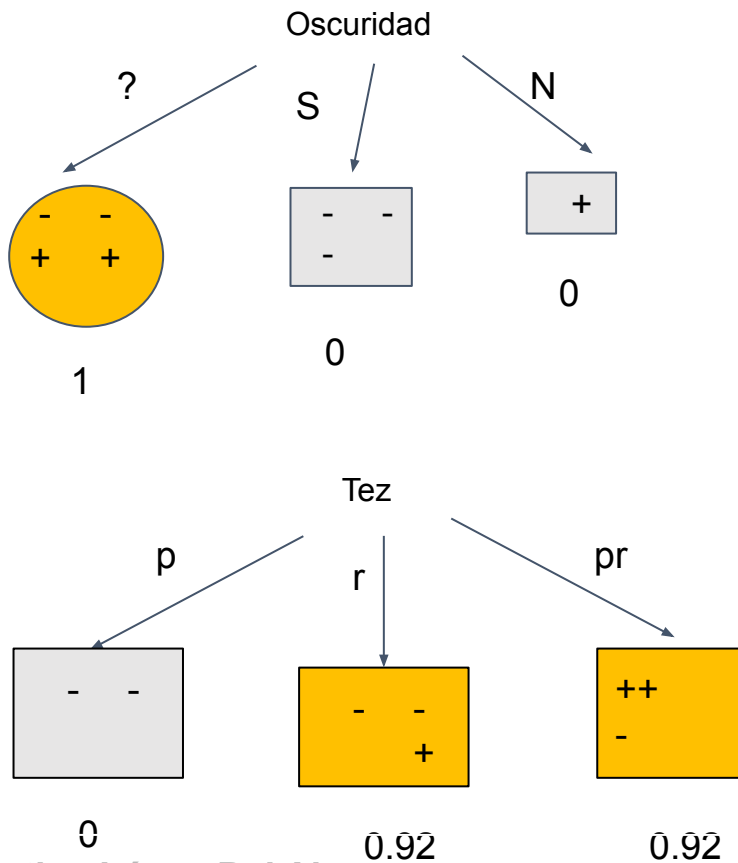
¿Cuál es el nivel de desorden de las 3 clases en Ajo?



$$D(CN) =$$

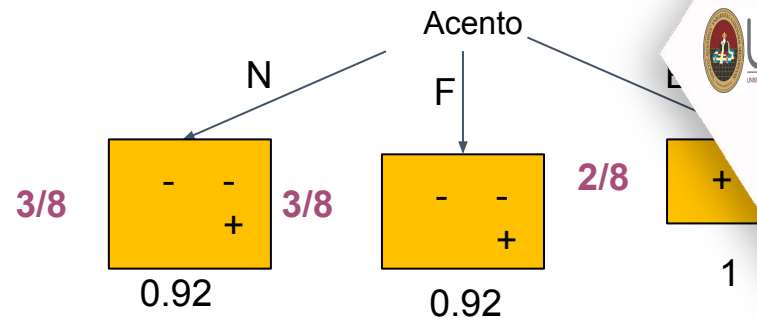
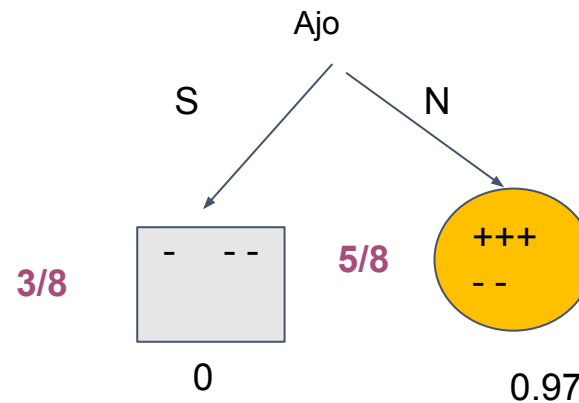
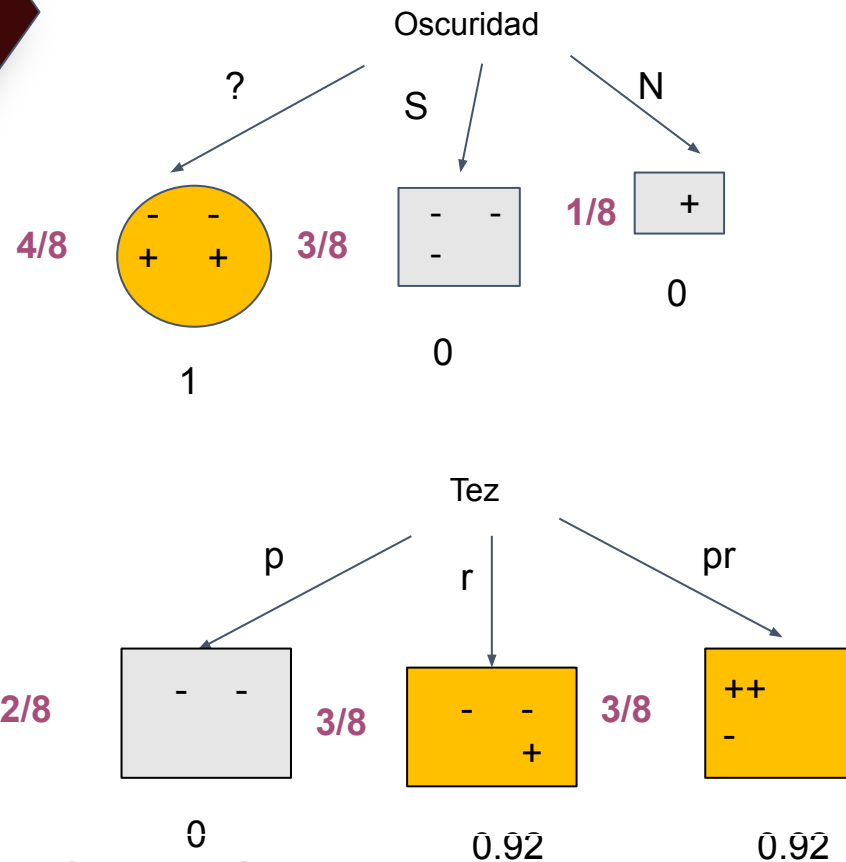
$$D(CF) =$$

$$D(CE) =$$

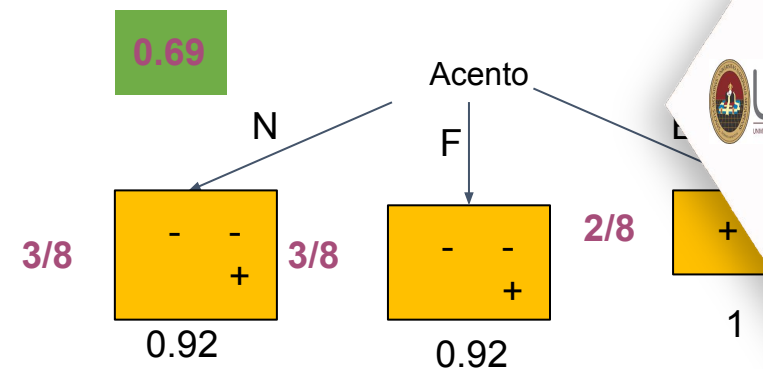
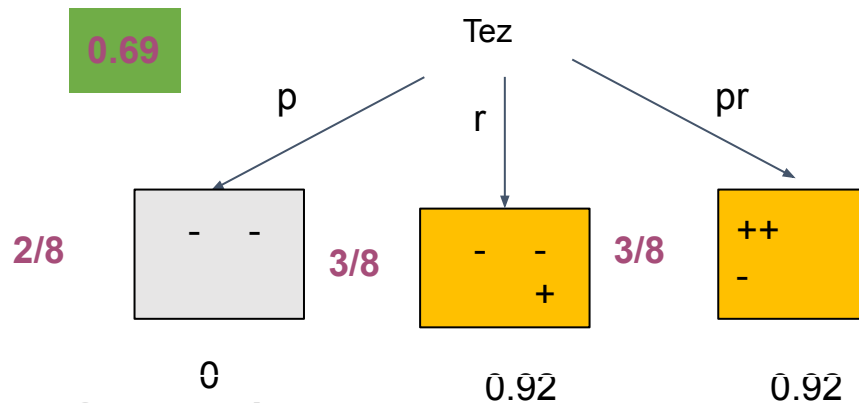
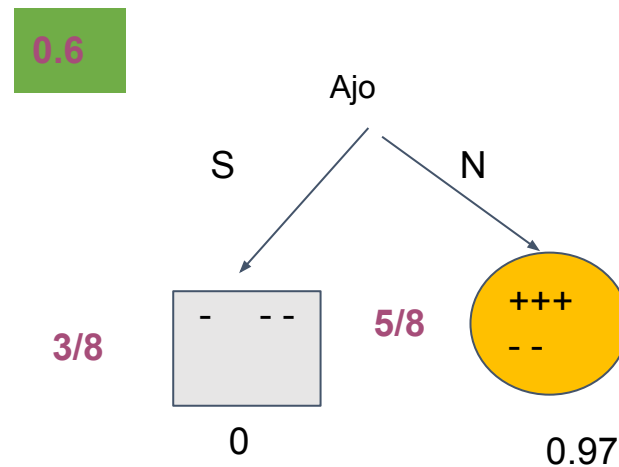
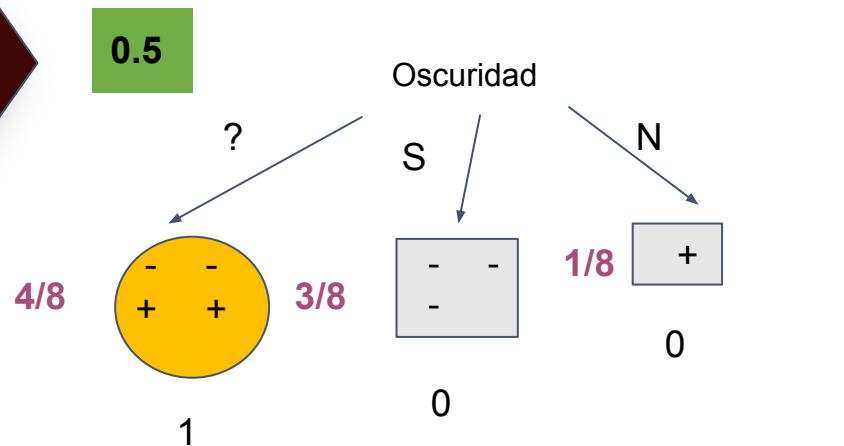




# Pesos de cada clase



# Pesos de cada clase



# 3

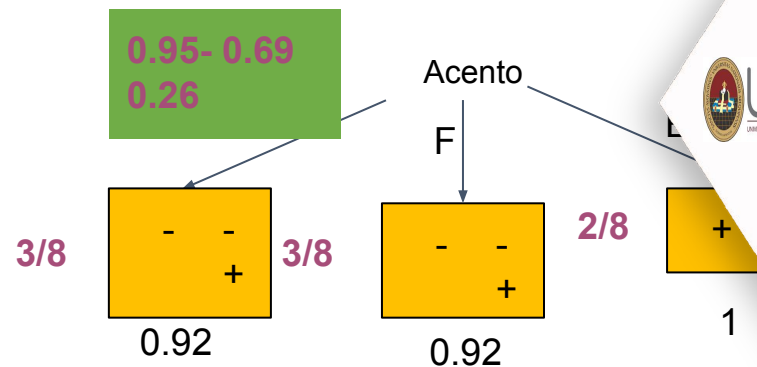
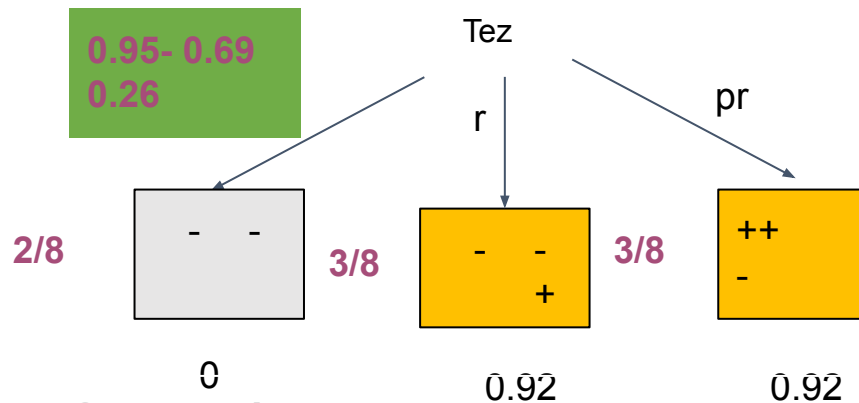
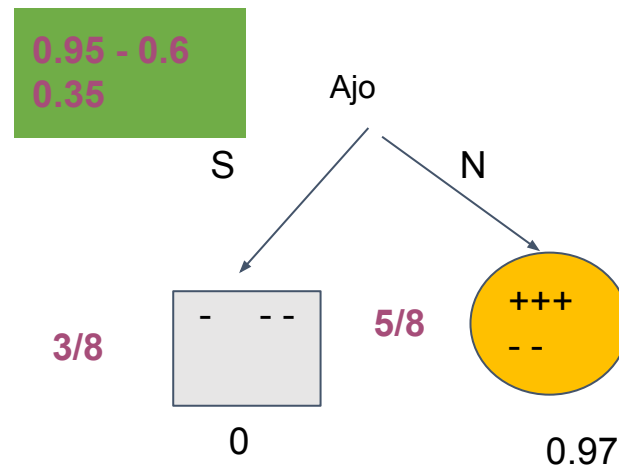
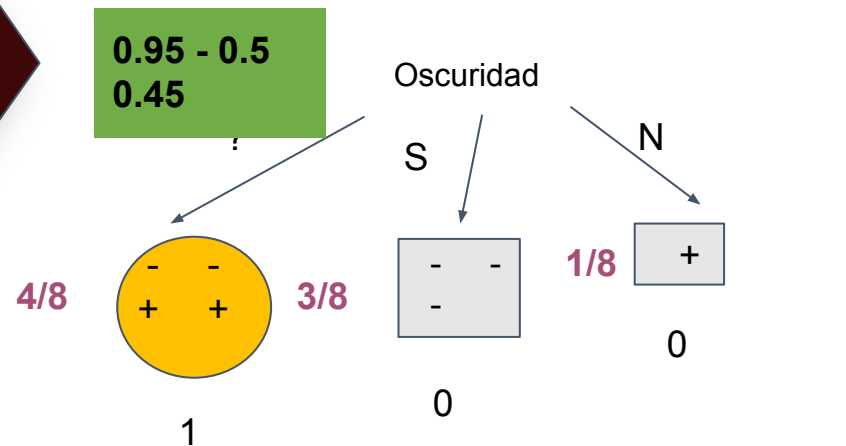
## Information Gain

## Ganancia de Información en cada Feacture

$$Gain(S) = D_s(S) - \sum_{f \in Feactures} \frac{|S_f|}{|S|} * D_s(S_f)$$

$$D_s(S) = -\frac{5}{8} \log_2 \frac{5}{8} - \frac{3}{8} \log_2 \frac{3}{8} = 0.954$$

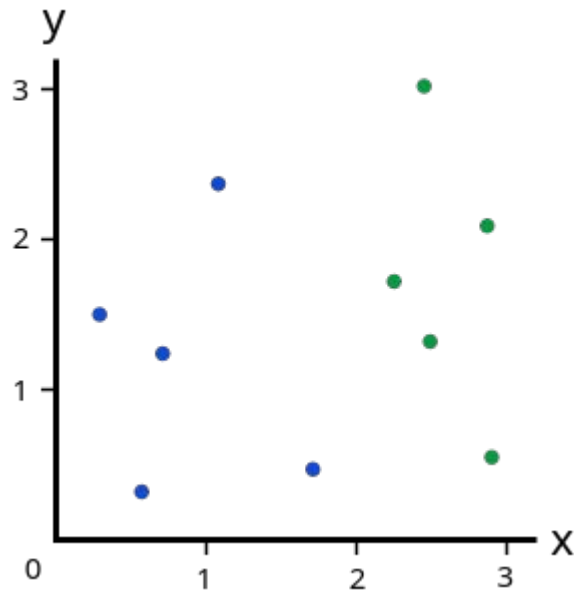
# Pesos de cada clase



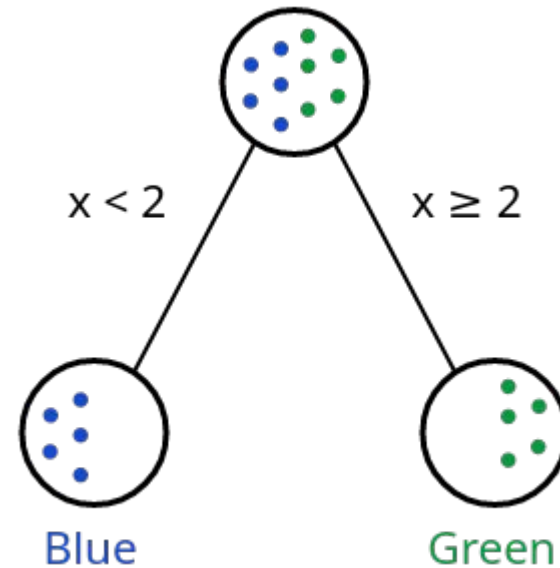
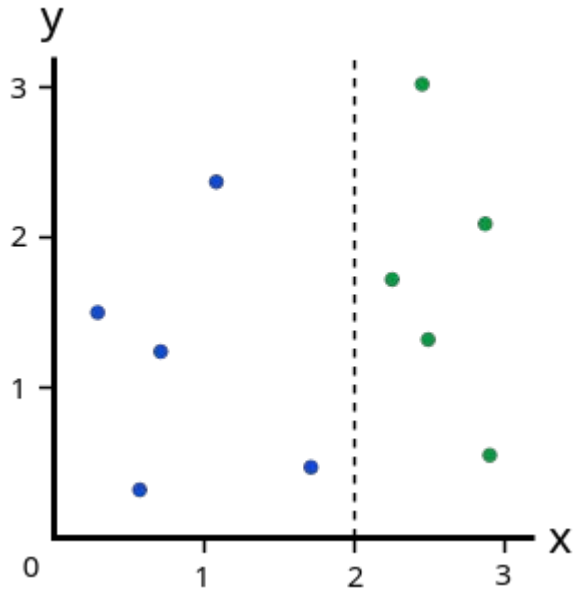
## Function ID3

- **Input:** Example set  $S$
- **Output:** Decision Tree  $DT$
- If all examples in  $S$  belong to the same class  $c$ 
  - return a new leaf and label it with  $c$
- Else
  - i. Select an attribute  $A$  according to some heuristic function
  - ii. Generate a new node  $DT$  with  $A$  as test
  - iii. For each Value  $v_i$  of  $A$ 
    - (a) Let  $S_i =$  all examples in  $S$  with  $A = v_i$
    - (b) Use ID3 to construct a decision tree  $DT_i$  for example set  $S_i$
    - (c) Generate an edge that connects  $DT$  and  $DT_i$

## Árbol de Decisión: Regresión

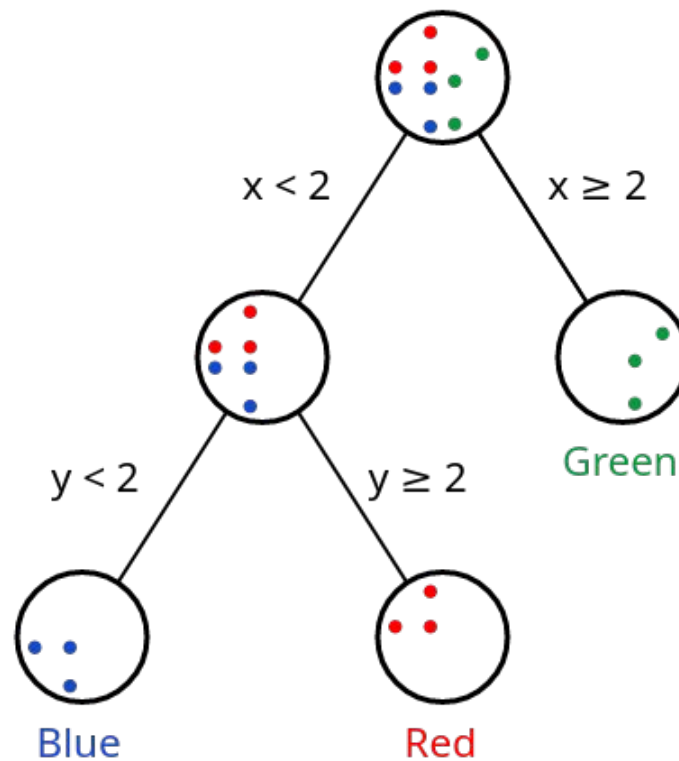
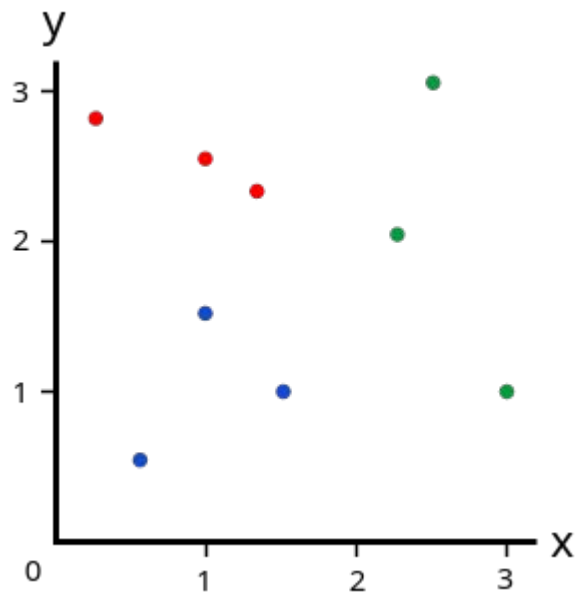


## Árbol de Decisión

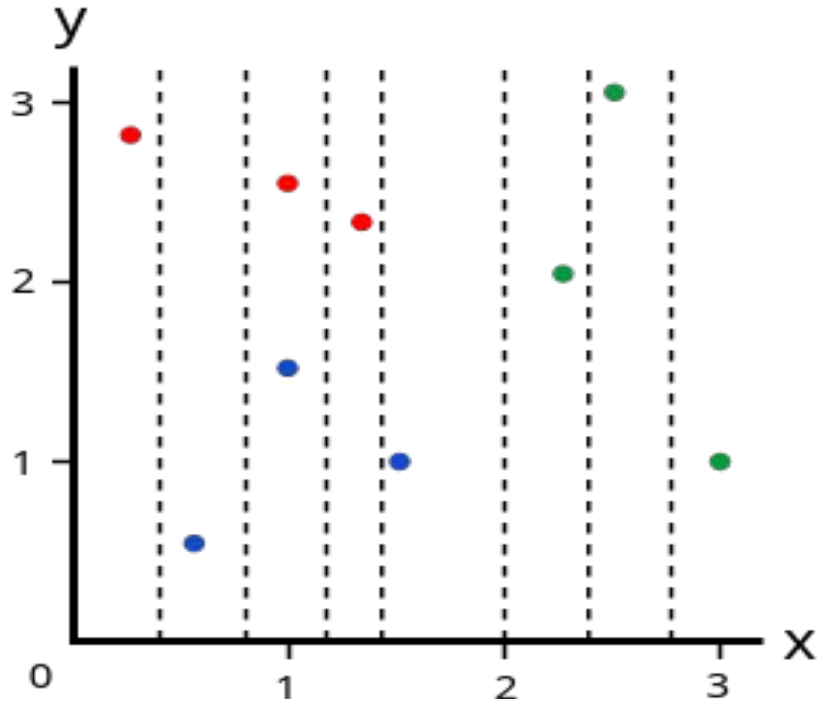




## Árbol de Decisión



## Árbol de Decisión



¿Cómo puedes el mejor split?



---

**Gracias**

