

Intuición  
febrero de 2022

(Hitz Research Center), Departamento de LSI, EHU

# Sist. de Ayuda a la Decisión:

## Clasificación: Árboles de Decisión

AitZiber AtutXa  
aitziber.atucha@ehu.eus

febrero de 2022

- 1 Objetivos de Aprendizaje
- 2 Motivación
- 3 Intuición tras los Árboles de Decisión
- 4 Construcción de los Árboles de Decisión
- 5 Formalización de los Árboles de Decisión
- 6 Lectura y visionado de vídeos

# Objetivos de Aprendizaje

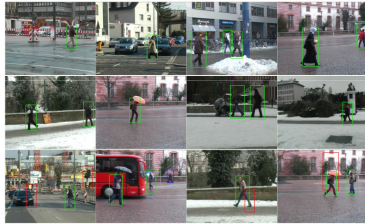
- Ser capaces de explicar la intuición tras el algoritmo.
- Ser capaces de formalizar matemáticamente el algoritmo.
- Ser capaces de contruir un Arbol de Decisión empleando la Ganancia de Información o empleando Gini.
- Ser capaces de generar un modelo básico de clasificación basado en el algoritmo de Arboles de decisión.

En el ámbito de los negocios

- Emplear datos históricos de ventas para predecir las consecuencias de decisiones
- Emplear datos demográficos para encontrar posibles clientes
- Emplear datos históricos para predecir qué clientes nos van a abandonar

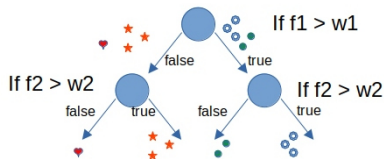
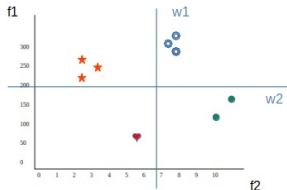
En otros ámbitos:

- En ML para detección de objetos.
- En juegos de interacción. Kinect (2011-2017)



# Intuición tras los Árboles de Decisión

Tomando decisiones con respecto a los atributos puedo llegar a caracterizar los datos

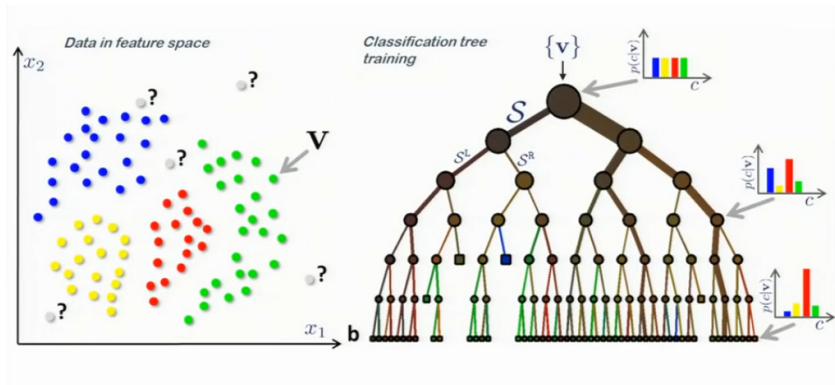


Pregunta: ¿Cuál es la probabilidad.

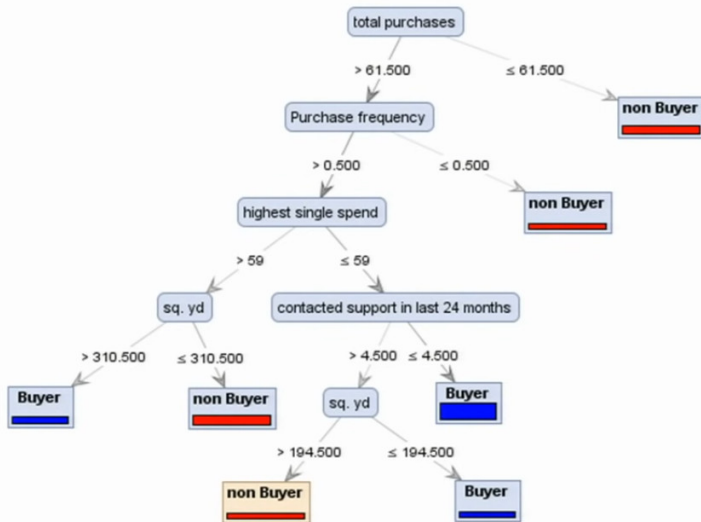
- a priori de estrella?
- y después del primer nodo (false)? y del primer nodo (true)?



# Intuición tras los Árboles de Decisión



# Intuición tras los Árboles de Decisión





# Construyendo el Arbol

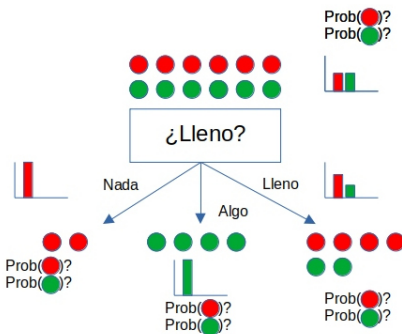
## Ejercicio: ¿Podemos permitirnos hacer esperar al cliente?

	Alt	Bar	Vier	Hamb	Lleno	Prec	Lluv	Res	Tipo	Estim	HacerEsp
$x^{1\}$	Si	No	No	Si	Algo	\$\$\$	No	Si	Frances	0-10	Si
$x^{2\}$	Si	No	No	Si	Lleno	\$	No	No	Thai	30-60	No
$x^{3\}$	No	Si	No	No	Algo	\$	No	No	Burger	0-10	Si
$x^{4\}$	Si	No	Si	Si	Lleno	\$	No	No	Thai	10-30	Si
$x^{5\}$	Si	No	Si	No	Lleno	\$\$\$	No	Si	Frances	>60	No
$x^{6\}$	No	Si	No	Si	Algo	\$\$\$	Si	Si	Italiano	0-10	Si
$x^{7\}$	No	Si	No	No	Nada	\$	Si	No	Burger	0-10	No
$x^{8\}$	No	No	No	Si	Algo	\$\$	Si	Si	Thai	0-10	Si
$x^{9\}$	No	Si	Si	No	Lleno	\$	Si	No	Burger	>60	No
$x^{10\}$	Si	Si	Si	Si	Lleno	\$\$\$	No	Si	Italiano	10-30	No
$x^{11\}$	No	No	No	No	Nada	\$	No	No	Thai	0-10	No
$x^{12\}$	Si	Si	Si	Si	Lleno	\$	No	No	Burger	30-60	Si



# Construyendo el Arbol

	Lleno	HacerEsp
$x^{\{1\}}$	Algo	Si
$x^{\{2\}}$	Lleno	No
$x^{\{3\}}$	Algo	Si
$x^{\{4\}}$	Lleno	Si
$x^{\{5\}}$	Lleno	No
$x^{\{6\}}$	Algo	Si
$x^{\{7\}}$	Nada	No
$x^{\{8\}}$	Algo	Si
$x^{\{9\}}$	Lleno	No
$x^{\{10\}}$	Lleno	No
$x^{\{11\}}$	Nada	No
$x^{\{12\}}$	Lleno	Si

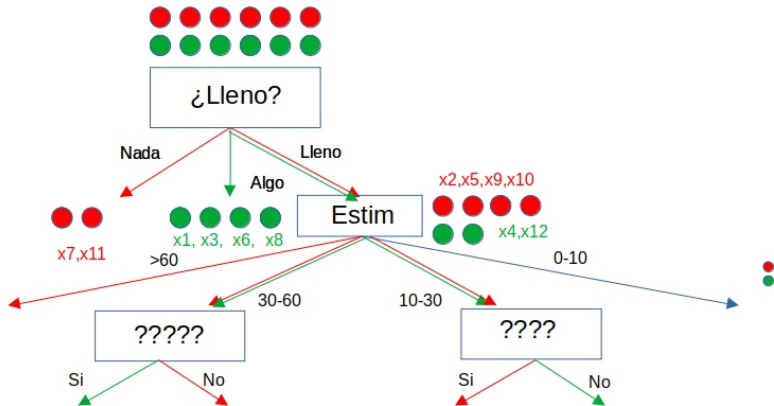


Pregunta: ¿Necesita este tipo de algoritmo...

que los atributos sean numéricos? (aplicar *cat2num*)  
escalar? (p.e. z-score) discretizar?

□

# Construyendo el Arbol



Pregunta: ¿Qué atributo seleccionaremos como root (raiz)?

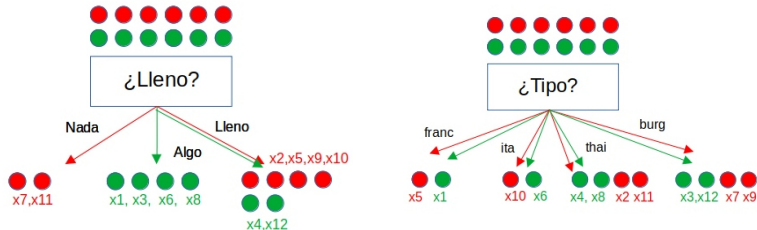
¿Cuál divide de forma más *discriminativa* los datos?

¿Cuál obtiene subconjuntos menos heterogéneos?

¿Qué utilizo como medida de discriminación?



# Construyendo el Arbol



Pregunta: ¿Cuál aporta más información?

¿Cuál es más discriminativo?

¿Cuál consigue subconjuntos menos heterogéneos?



## Entropía

La entropía mide la heterogeneidad de un a muestra.  
Cuanto mas heterogénea más entropía.

$$H(\pi) = - \sum \pi \log(\pi)$$

- Clasificación binaria:

$$H\left(\frac{pos}{pos+neg}, \frac{neg}{pos+neg}\right) = -\frac{pos}{pos+neg} \log_2 \frac{pos}{pos+neg} - \frac{neg}{pos+neg} \log_2 \frac{neg}{pos+neg}$$

- Clasificación multiclase:

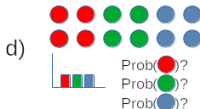
$$H(P(c_1), \dots, P(c_n)) = - \sum_{c=1}^n P(c_i) \log(P(c_i))$$

# Construyendo Arbol: Intuición de Entropía

Pregunta: ¿Cuáles son más heterogéneos?

Ordenalos de mayor a menor (empates iguales) y calcula la Entropía de cada uno (Sol: Después)

□



# Const. Arbol: Entropía estim. y ganancia de inf.

Ahora ya tengo todo lo que necesito:

- Tengo los datos (grupo de instancias) antes de tomar la decisión
- Tengo las posibles decisiones (atributos)
- Tengo una medida que calcula lo heterogéneo que es un grupo de instancias

Pregunta: ¿tiene sentido...

- 1 calcular la heterogeneidad antes de aplicar un atributo como decisor
- 2 calcular la estimación de heterogeneidad después de aplicar un atributo como decisor
- 3 considerar la resta de ambos como ganancia de información o tasa de discriminación del decisor?





# Const. Arbol: Entropía estim. y ganancia de inf.

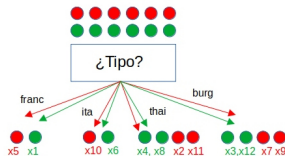
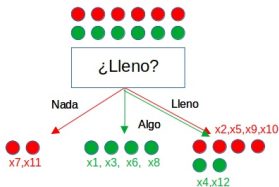
atributos-decisores (features):  $F = \{f_1, f_2, \dots, f_n\}$

ramas del decisor:  $V = \{v_1, v_2, \dots, v_n\}$

## Entropía Estimada y Ganancia de Información

$$EH(f_j) = \sum_{i=1}^V \frac{p_i + n_i}{p_{padre} + n_{padre}} H\left(\frac{p_i}{(p_i + n_i)}, \frac{n_i}{(p_i + n_i)}\right)$$

$$IG(f) = H\left(\frac{p_{padre}}{p_{padre} + n_{padre}}, \frac{n_{padre}}{p_{padre} + n_{padre}}\right) - EH(f)$$



# Formalización de los Árboles de Decisión

Recordemos:  $\mathcal{D}^{train}$ ,  $\mathcal{D}^{dev}$ ,  $\mathcal{D}^{test}$

$$\mathcal{X}^{train} = \{x^{\{1\}}, x^{\{2\}}, \dots, x^{\{n\}}\}, \mathcal{X}^{dev} = \{x^{\{n+1\}}, x^{\{n+2\}}, \dots, x^{\{l\}}\},$$

$$\mathcal{X}^{test} = \{x^{\{l+1\}}, x^{\{l+2\}}, \dots, x^{\{z\}}\}$$

$n$  es tamaño de la muestra train,  $l-n+1$  el tamaño del dev,  $z-l-n+1$   
(i.e. núm. instancias de entrenamiento, de desarrollo, de test)

$$\mathcal{Y}^{train} = \{y^{d\{1\}}, y^{d\{2\}}, \dots, y^{d\{n\}}\}, \mathcal{Y}^{dev} = \{y^{d\{n+1\}}, y^{d\{n+2\}}, \dots, y^{d\{l\}}\}$$

$$\mathcal{Y}^{test} = \{y^{d\{l+1\}}, y^{d\{l+2\}}, \dots, y^{d\{z\}}\}$$

Las  $\mathcal{Y}$  contienen los valores reales a acertar

Obviamente  $\mathcal{Y}^{train}$  contiene  $n$  valores porque cada instancia de  $\mathcal{X}^{train}$  tendrá su clase real asociada, lo mismo para el dev y test

$$\mathcal{D}^{train} = \{(x^{\{1\}}, y^{d\{1\}}), (x^{\{2\}}, y^{d\{2\}}), \dots, (x^{\{n\}}, y^{d\{n\}})\}$$

$$\mathcal{D}^{dev} = \{(x^{\{n+1\}}, y^{d\{n+1\}}), (x^{\{n+2\}}, y^{d\{n+2\}}), \dots, (x^{\{l\}}, y^{d\{l\}})\}$$

$$\mathcal{D}^{test} = \{(x^{\{l+1\}}, y^{d\{l+1\}}), (x^{\{l+2\}}, y^{d\{l+2\}}), \dots, (x^{\{z\}}, y^{d\{z\}})\}$$

$$\text{Idealmente } \mathcal{D}^{train} \cap \mathcal{D}^{dev} \cap \mathcal{D}^{test} = \emptyset$$

# Formalización de los Árboles de Decisión

$$F = f_1, f_2, \dots, f_n$$

---

## Algorithm 1 ArbolDec( $\mathcal{D}$ , $F$ )

---

```
Pred = la clases más frecuente en  $\mathcal{D}$   
if todas clases  $\in \mathcal{D}$  son iguales then  
    return Pred  
else if  $F$  esta vacío then  
    return Pred  
else if  $\mathcal{D}$  esta vacío then  
    Pred =  $\arg \max(\text{Prob}(\mathcal{D}\{\text{train}\}|f))$  return Pred  
else  
    for each  $f_i \in F$  do  
         $NO \subseteq \mathcal{D}|f_i = \text{false}$   
         $SI \subseteq \mathcal{D}|f_i = \text{true}$   
         $\text{score}(f_i) = \text{GananciaDeInf}(f_i)$   
    end for  
     $f_{\max} = \arg \max(\text{score}(F))$   
     $NO \subseteq \mathcal{D}|f_{\max} = \text{false}$   
     $SI \subseteq \mathcal{D}|f_{\max} = \text{false}$   
     $F = F - f_{\max}$   
     $\text{izq} = \text{ArbolDec}(NO, F)$   
     $\text{drch} = \text{ArbolDec}(SI, F)$   
    return Nodo( $f, \text{izd}, \text{drch}$ )  
end if
```

## Ganancia de Inf. vs. Gini

**Desvent:** Tienden a sobreajustar (overfitting). Es decir, generalizan poco.

- Pequeños cambios en la muestra = gran impacto en predic (high variance algorithm).
- Posibles soluciones:
  - Poda (imponer un max. profundidad)
  - min. num. instancias
  - Threshold de IG ..

**Vent:** Altamente interpretables

**Vent:** Preprocesado min. (no cat2num, no escalado...)

**Vent:** No hace falta una muestra muy grande

## Ganancia de Inf. vs. Gini

Ganancia de Inf.

$$IG(f) = H\left(\frac{p_{padre}}{p_{padre}+n_{padre}}, \frac{n_{padre}}{p_{padre}+n_{padre}}\right) - EH(f)$$

Gini

$$Gini(f_i) = 1 - \sum_{i=1}^n Prob(i)^2$$

Ejercicio: Elegir la raíz aplicando Gini ...

$Gini(f_i = Lleno)$

$Gini(f_i = Tipo)$



- "Artificial Intelligence: A Modern Approach", eBook, Global Edition (English Edition) S. Russell P. Norvig Pearson Chap 18.3 <https://cs.calvin.edu/courses/cs/344/kvlinden/resources/AIMA-3rd-edition.pdf>
- Vídeo con un ejemplo de cómo aplicar el índice Gini <https://es.coursera.org/lecture/build-decision-trees-svms-neural-networks/gini-index-example-rPvWM>

- "Artificial Intelligence: A Modern Approach", eBook, Global Edition (English Edition) S. Russell P.Norvig Pearson Chap 18.3 <https://cs.calvin.edu/courses/cs/344/kvlinden/resources/AIMA-3rd-edition.pdf>