

T5-aa.pdf



Olmar_eps



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.

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

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

Tema 5: APRENDIZAJE AUTOMÁTICO

5.1. INTRODUCCIÓN a ML.

■ CONCEPTOS GENERALES

- > **Atributo:** característica que describe parcialmente a los elementos (columna de una tabla)
- > **Instancia:** un elemento definido por los valores de sus atributos (fila de una tabla)
- > **Clase:** subconjuntos disjuntos (categorías) en los que se quiere dividir el conjunto de instancias \rightarrow una de las columnas
 - \hookrightarrow Atributo a predecir.
- Ejemplo \oplus de una clase: instancia que pertenece al subconj. def. por la clase.
- Ejemplo \ominus de una clase: instancia que NO pertenece al subconj. def. por la clase.

■ PRE-PROCESADO

- Preparar datos para modelar
 - \rightarrow Limpieza: si le faltan valores (NaN), también debe preprocesar.
 - \rightarrow Outliers: puntos de datos que difieren significativamente de otras observaciones
 - \rightarrow El modelado en ML requiere entradas numéricas.
- Si el dataset NO es suficientemente grande $\left\{ \begin{array}{l} \text{No podemos permitirnos perder algunas filas} \\ \text{Aproximamos el valor de los datos faltantes} \end{array} \right.$
- Normalización logarítmica: usar en columnas con alta varianza.

■ MÉTODOS

- \hookrightarrow Árbol de decisión (CART)
- \hookrightarrow Aprendizaje basado en instancias: **K-NN**

$d = \sqrt{\sum (x_i - y_i)^2}$

Dada una observación x se buscan las K observaciones más cercanas a x (K -vecinos) entre los x_i del conjunto de entrenamiento. Se asigna x a la clase mayoritaria entre los vecinos.
- \hookrightarrow Aprendizaje Bayesiano: **red bayesiana**

Las decisiones se toman en función de los datos observados

 - Clasificador Bayesiano $(x_i) = \arg\max P(x_i | C_j) P(C_j)$
 - Naive Bayes $(x) = \arg\max P(C_j) \prod P(x_i | C_j) \rightarrow$ Red NB



WUOLAH

5.2. ÁRBOLES de DECISIÓN

- Particiona el espacio de la muestra en rectángulos y luego predice un modelo simple en cada uno de ellos
- Los árboles binarios van discriminando el espacio en dos submuestras (nodos) a partir del anterior.



9/1/2
1

OBJETIVO

- Construir el árbol más sencillo que mejor separe las instancias por clase
- El modelo final debe generalizar para clasificar bien futuras instancias.

¿Cómo funciona?

- La clase asociada a un nodo corresponde a la etiqueta mayoritaria de las instancias de entrenamiento asignadas a ese nodo
- Los tests en los nodos internos se determinan maximizando una cantidad que favorece una separación más clara de las clases en los hijos de dicho nodo.

Criterios de separación:

- Entropía → mide la ausencia de homogeneidad de un conjunto de ejemplos respecto a su clase.
 - Impureza de Gini
 - Garancia de información → diferencia entre entropía conj. original y subconj. obtenidos
 - Ratio de garancia de información.
 - Precisión
- El proceso se detiene cuando añadir un nuevo atributo no mejora el criterio de separación.
 - La predicción de etiqueta de clase tiene lugar en el nodo hoja.

■ ID3

En cada iteración:

- ↳ Obtener el atributo en base al cual ramificar el nodo problema
- ↳ Seleccionar el que mejor discrimine entre el conjunto de ejemplos
 - * Heurística para obtener árboles pequeños (en profundidad)
 - * El atributo más discriminante será aquel que conduzca a un estado con menor entropía o menor desorden (mayor información)

Para cada atributo se calcula la disminución de entropía:

$$\text{Disminución Entropía}_A = E(X) - E_A(X)$$

En cada nodo, se selecciona aquel atributo que mayor disminución de entropía proporcione

Esta medida tiende a favorecer la elección de $\left\{ \begin{array}{l} \text{atributos con muchos valores posibles} \\ \text{lo que redunde en una peor generaliz.} \end{array} \right.$

(2)

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.

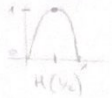
0

Sin necesidad de domiciliar nómina o recibos.

▷ ENTROPÍA BINARIA

$$H(x) = H_b(p) = -p \log_2(p) - q \log_2(q) \text{ [bits]}$$

$$0 \leq H(x) \leq 1$$



▷ ENTROPÍA V.A. DISCRETA

→ mide la cantidad promedio de bits por símbolo del mensaje de longitud mínima

$$H(x) = - \sum_i P(x=x_i) \log_2 P(x=x_i) \rightarrow H(3+, 3-) = 1 \Rightarrow 0 \text{ información}$$

▷ IMPUREZA de GINI BINARIA

$$I_G(x) = 1 - p^2 - q^2 \rightarrow \text{parecida a entropía}$$

▷ IMPUREZA de GINI V.A. DISCRETA

$$I_G(x) = 1 - \sum_i [P(x=x_i)]^2$$

▷ ENTROPÍA CONDICIONAL

$$H(Y|X) = - \sum P(x=x_i) H(Y|X=x_i)$$

$$H(Y|X) \leq H(Y)$$

$$H(Y|X) = H(Y) \text{ si } X \text{ e } Y \text{ indep.}$$

▷ GANANCIA de INFORMACIÓN

→ mide n° promedio de bits por símbolo del mensaje mínimo

$$IG(Y|X) = H(Y) - H(Y|X) \text{ [bits]}$$

Selecciona el mejor atributo al que maximiza la IG de la clase

→ mejor atributo para hacer división en árbol raíz

ID3: Se elige atributo!

→ Partición Atributo.

$$\begin{aligned} & \text{Ej. } \begin{matrix} \text{BIN} \\ \text{Atrib} \end{matrix} \begin{matrix} \text{VALORES} \\ P(A) \\ P(B) \\ P(C) \end{matrix} \begin{matrix} \text{Ej.} \\ P(N) \text{ n° sí, n° no} \\ P(SU) \text{ n° sí, n° no} \\ P(B) \text{ n° sí, n° no} \end{matrix} \\ & \begin{matrix} \text{① } P(\text{Atr Yes}) \\ P(\text{Atr No}) \end{matrix} \begin{matrix} \text{② } H(\text{cl}|\text{Atr Yes}) \\ H(\text{cl}|\text{Atr No}) \end{matrix} \begin{matrix} \text{③ } H(\text{cl}|\text{Atr}) \\ I_G(\text{cl}|\text{Atr}) = H(\text{cl}) - H(\text{cl}|\text{Atr}) \end{matrix} \end{aligned}$$

C4.5: Se elige Pregunta! → aunque sea binaria

→ En cada valor del atributo (si no es binario) → y luego a esos valores

$$\begin{aligned} & \text{Ej. } \begin{matrix} \text{BIN} \\ \text{Pregunta} \end{matrix} \begin{matrix} \text{VALORES} \\ P(\text{Valor 1}) \\ P(\text{Valor 2}) \end{matrix} \begin{matrix} \text{Ej.} \\ P(N) \text{ n° sí, n° no} \\ P(SU) \text{ n° sí, n° no} \\ P(B) \text{ n° sí, n° no} \end{matrix} \\ & \begin{matrix} \text{① } P(\text{Valor 1}) \\ P(\text{Valor 2}) \end{matrix} \begin{matrix} \text{② } H(\text{cl}|\text{Valor 1}) \\ H(\text{cl}|\text{Valor 2}) \end{matrix} \begin{matrix} \text{③ } H(\text{cl}|\text{Valor}) \\ I_G(\text{cl}|\text{Valor}) = H(\text{cl}) - H(\text{cl}|\text{Valor}) \end{matrix} \end{aligned}$$

$$\begin{aligned} & \text{Ej. } \begin{matrix} \text{BIN} \\ \text{Pregunta} \end{matrix} \begin{matrix} \text{VALORES} \\ P(\text{Valor 1}) \\ P(\text{Valor 2}) \end{matrix} \begin{matrix} \text{Ej.} \\ P(N) \text{ n° sí, n° no} \\ P(SU) \text{ n° sí, n° no} \\ P(B) \text{ n° sí, n° no} \end{matrix} \\ & \begin{matrix} \text{① } P(\text{Valor 1}) \\ P(\text{Valor 2}) \end{matrix} \begin{matrix} \text{② } H(\text{cl}|\text{Valor 1}) \\ H(\text{cl}|\text{Valor 2}) \end{matrix} \begin{matrix} \text{③ } H(\text{cl}|\text{Valor}) \\ I_G(\text{cl}|\text{Valor}) = H(\text{cl}) - H(\text{cl}|\text{Valor}) \end{matrix} \end{aligned}$$



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

UNDERFITTING / OVERFITTING

	Sesgo	Varianza
MODELO SIMPLE ↳ underfitting	ALTO ↑	PEQUEÑA ↓
MODELO COMPLEJO ↳ overfitting	BAJO ↓	GRANDE ↑

Idea: árboles grandes y reducir complejidad → POST-PODA

¿Cuándo se deja de dividir un nodo?

- Los ejemplos de entrenamiento asignados a ese nodo pertenecen a la misma clase.
- El nodo no tiene ejemplos asociados.
- No quedan más atributos para dividir los datos.
- Poda previa (con límite = tam. árbol para evitar sobreajuste) $\leftarrow G1$
nº ejemplos por debajo umbral.

C4.5

Soluciona problema de ID3 ya que ID3 tiende a favorecer la elección de atributos con muchos valores posibles, lo que redundaría en una peor generalización de los obs.

→ Usa Ratio de Ganancia para cada atributo.

$$H(\text{clase} | \text{Pregunta}) = R.12Q * H(12Q) + R.DCHA * H(DCHA)$$

$$IG = H(\text{clase}) - H(\text{clase} | \text{Pregunta})$$

	PARÁMETROS (determinados en entrenamiento)	HIPERPARÁMETROS (forma en la que se determina el entrenamiento) no resto
Red Neuronal	PESOS SINÁPTICOS	CTE APRENDIZAJE / MAGNITUD REGULAR ↳ + peque ⇒ + tarda en aprender ↳ + alto ⇒ función mal regular en ese log
Árboles de Decisión	ATRIBUTOS + UMBRALES ETIQU. CLASES	PROFUNDIDAD MÁXIMA (nº min de q's)
K-NN	$D_{\text{train}} = \{x_n, y_n\}$	k, distancia

5.3. REDES NEURONALES

■ PERCEPTRÓN MONOCAPA

→ Regresión logística.

$$w^* = \underset{w}{\operatorname{argmin}} E(w)$$

• Error de entropía cruzada:

$$E(w) = - \sum_{n=1}^N ((1-t_n) \log(1-\sigma(w^T x_n)) + t_n \log \sigma(w^T x_n))$$

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

• Gradiente de entropía cruzada

$$\frac{\partial}{\partial w} E(w) = \sum_{n=1}^N (\sigma(w^T x_n) - t_n) x_n = \sum_{n=1}^N \delta_n x_n$$

$$\delta_n = \sigma(w^T x_n) - t_n$$

→ Aprendizaje por lotes (BATCH LEARNING)

- w's iguales en todas las filas
- Solo para final actualizo los pesos:

$$w_f = w - \eta \sum_n \delta_n x_n = w + \sum \Delta w_{in}$$

$$\Delta w_i = -\delta_i \cdot \eta \cdot x_i$$

$$z = w^T x_n, \quad \delta_n = \sigma(z_n) - t_n, \quad \sigma(z) = \frac{1}{1+e^{-z}}$$

\parallel
 $w_0 + w_1 x_1 + \dots + w_n x_n$

→ Aprendizaje en línea (ONLINE)

- cambio w's en cada paso → actualizo los pesos en cada fila
- En cada paso

$$w_{i,t+1} = w_i - \eta \delta_i x_{i,t} = w_i + \Delta w_{i,t}$$

$$\Delta w_i = -\delta_i \cdot \eta \cdot x_i$$

$$z = w^T x_n, \quad \delta_n = \sigma(z_n) - t_n, \quad \sigma(z) = \frac{1}{1+e^{-z}}$$

CODIFICACIÓN NUMÉRICA: (atributos)

→ K vecinos próximos → para calcular distancias

→ Redes neuronales → Pz procesan datos numéricos

NO → NB → puede trabajar directamente sobre datos cualitativos

NO → Modelo gráfico → " " " " " " " " " " " "