

Nociones previas a las prácticas de

Técnicas de Aprendizaje Automático

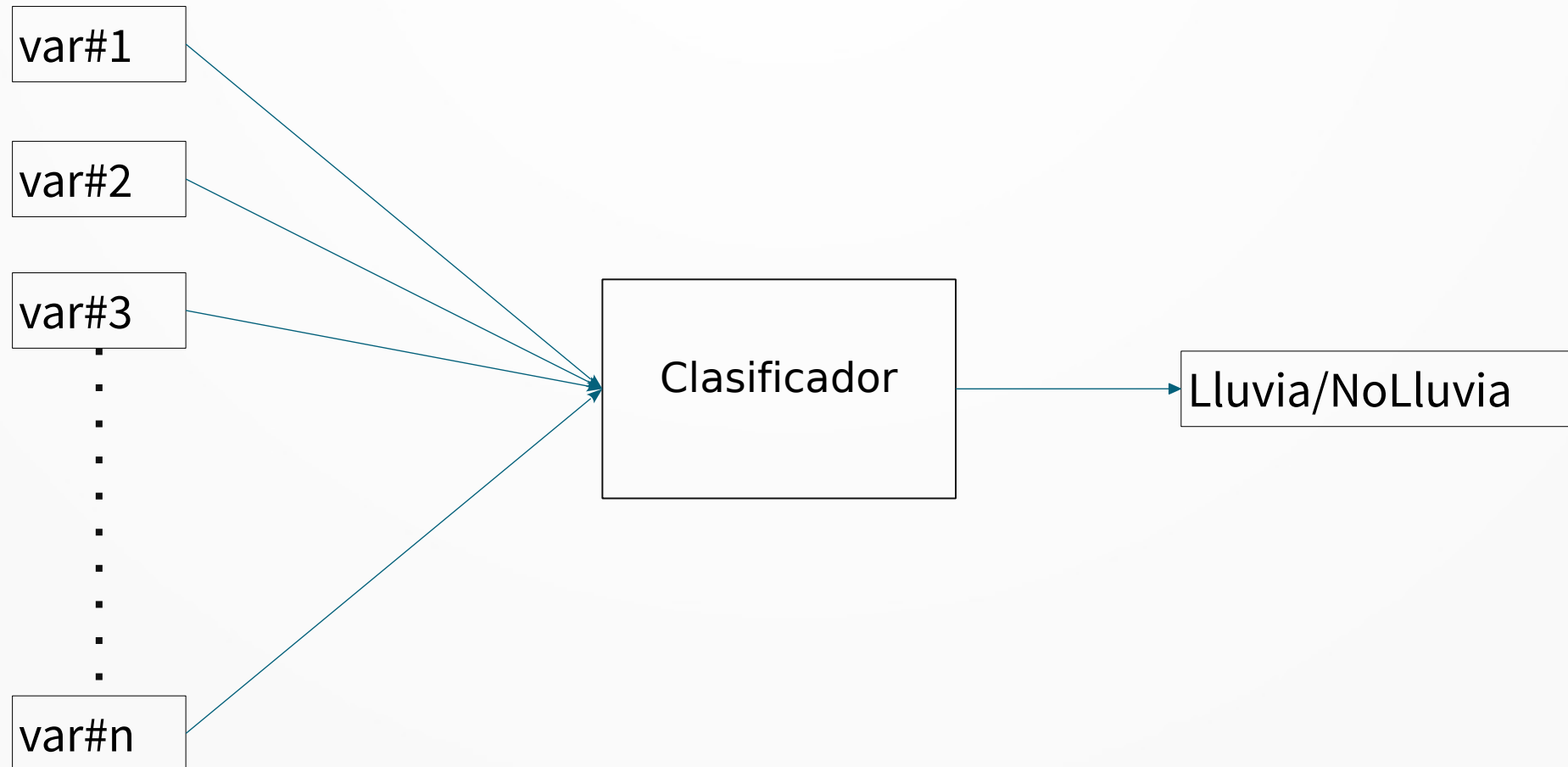


UNIVERSIDAD DE VALLADOLID

Técnicas de Aprendizaje Automático
Grado de Ingeniería Informática
Doble Grado con Estadística (INDAT)
Departamento de Informática (ATC, CCIA y LSI)

Esquema básico del Clasificador

Caso i-ésimo



Características del Clasificador

- Desde el punto de vista matemático: $f : \mathcal{R}^n \rightarrow R(\text{discreto})$
- La función f se obtiene por entrenamiento con **ejemplos**
- La calidad del clasificador se mide por la **generalización**
- Aunque es un tipo particular de aprendizaje (**supervisado**) es el más frecuente

Clasificación / Regresión

- La **base de conocimiento** se compone de ejemplos:

$$D = \{ \langle \bar{x}_1, \theta_1 \rangle, \langle \bar{x}_2, \theta_2 \rangle, \dots, \langle \bar{x}_N, \theta_N \rangle \}$$

donde \bar{x}_i es $\langle \text{var\#1}, \text{var\#2}, \dots, \text{var\#n} \rangle$ para cada uno de los N ejemplos. Puede haber mezcla de valores continuos o discretos.

Por su parte, θ_i es un valor **discreto** (Lluvia/No Lluvia)

- Cuando θ_i pasa a ser un valor continuo, se habla de **regresión**

Generalización

- Los experimentos organizan los ejemplos en **dos conjuntos**:
 - **Aprendizaje** (por ej: 2/3)
 - **Test** (por ej: 1/3)
- Ambos deben ser **disjuntos**.
- Esta técnica sencilla se denomina Hold-out o de **retención**.
- **Validación cruzada** (Tema de Metodología Experimental):
 - Se divide en conjunto original en K carpetas con idéntica distribución por clases
 - Se plantean k sub-experimentos, donde cada carpeta es usada de test y el resto de aprendizaje.

WEKA

- Se descarga de internet:



- <https://www.cs.waikato.ac.nz/ml/weka/>
- En Windows e IOS, hay un programa de **auto-instalación**
- En Linux hay que descargar un paquete y descomprimirlo en un directorio de trabajo. Hay que ejecutar un **script** (weka.sh)
- Se necesita Java Development Kit (**JDK**)
- Para su ejecución, en Windows y IOS se lanza con el **icono**.
- En Linux, se haría desde un terminal, o se construye un icono a través de un fichero .desktop, que debe ser alojado en:
 - /usr/share/applications/

- Atención al “**package manager**”, porque es posible que haya algoritmos que no están accesibles y haya que cargarlos con esta herramienta

Ejemplo en WEKA

- Cargar el fichero de datos **weather.nominal.arff**
- Construir un árbol de decisión usando el algoritmo **ID3**
 - Si no está accesible este algoritmo, habrá que cargarlo con el “package manager”
 - No se puede visualizar el árbol en modo gráfico, sí en formato texto plano
- Lo mismo, pero usando el algoritmo **J48**:
 - Sí se puede ver el árbol en modo gráfico

Python / Anaconda

- Se descarga de la red. Sólo la versión abierta (Python 3.7).
- A partir de ahora, con este software, sólo se usará Linux
- Desde un terminal, se activa el entorno de trabajo “**conda**”:
 - `$ conda activate`
- Para abrir una consola “python”, se empleará:
 - `$ jupyter-notebook`
 - `$ jupyter-lab`
- Existe una herramienta que permite ver todas las posibilidades de este paquete:
 - `$ anaconda-navigator`

Ejemplo Python

- Tomar el fichero de datos suministrado por “sklearn”:
 - **iris**
- Construir un árbol con el módulo de sklearn:
 - **DecisionTreeClassifier**
- Exportarlo a un fichero de comandos ‘**.dot**’
- Con la herramienta ‘dot’ del paquete Linux “**graphviz**”, crear un pdf con la imagen del árbol correspondiente