

Hipótesis o conceptos



MINERÍA DE DATOS

Reglas de Decisión no estadísticas

2

- Problema: lograr definir un concepto a partir de instancias rotuladas como pertenecientes y no pertenecientes al concepto.
- Instancias tienen características o atributos
- Hipótesis son las combinaciones posibles de valores de atributos.
- Muestras de entrenamiento validan las hipótesis posibles

Aprendizaje de conceptos

3

Ejemplo de aprendizaje de concepto o por inferencia inductiva:

- Concepto a aprender: *Días aptos para realizar un deporte de vela*
- *Instancias*: para cada día se tiene acceso a los siguientes atributos (con sus posibles valores):
 - ✦ Cielo: sol, nublado, lluvia
 - ✦ Temperatura del aire: cálido, frío
 - ✦ Humedad Ambiente: normal, alta
 - ✦ Viento: débil, fuerte
 - ✦ Temperatura del agua: caliente, fría
 - ✦ Pronostico: estable, cambio

Espacio de instancias: $3 \times 2 \times 2 \times 2 \times 2 \times 2 = 96$ posibilidades

- **Hipótesis**: *Conjunto de restricciones sobre los atributos*
Posibilidades para los atributos: \emptyset (no se permite), ? (cualquier cosa), valor

Aprendizaje de conceptos

4

- Muestras de entrenamiento en tabla Minería de Datos

Cielo	Aire	Humedad	Viento	Agua	Pronostico	Deporte
sol	cálido	normal	fuerte	caliente	estable	si
sol	cálido	alta	fuerte	caliente	estable	si
lluvia	frío	alta	fuerte	caliente	cambio	no
sol	cálido	alta	fuerte	fría	cambio	si

- Ejemplo de hipótesis: $\{\text{sol}, ?, \text{normal}, ?, ?, \text{cambio}\}$ (**F**)
- Hipótesis puede ser (**F**) o (**V**). Es (**F**) cuando no hay ninguna muestra que la cumpla en el conjunto de entrenamiento. (**V**) cuando hay al menos una.
- Hipótesis más específica: $\{\emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset\}$. *Resulta F* si el conjunto de entrenamiento es vacío.
- Hipótesis más general: $\{?, ?, ?, ?, ?, ?\}$ es **F** si el conjunto de entrenamiento no es vacío.

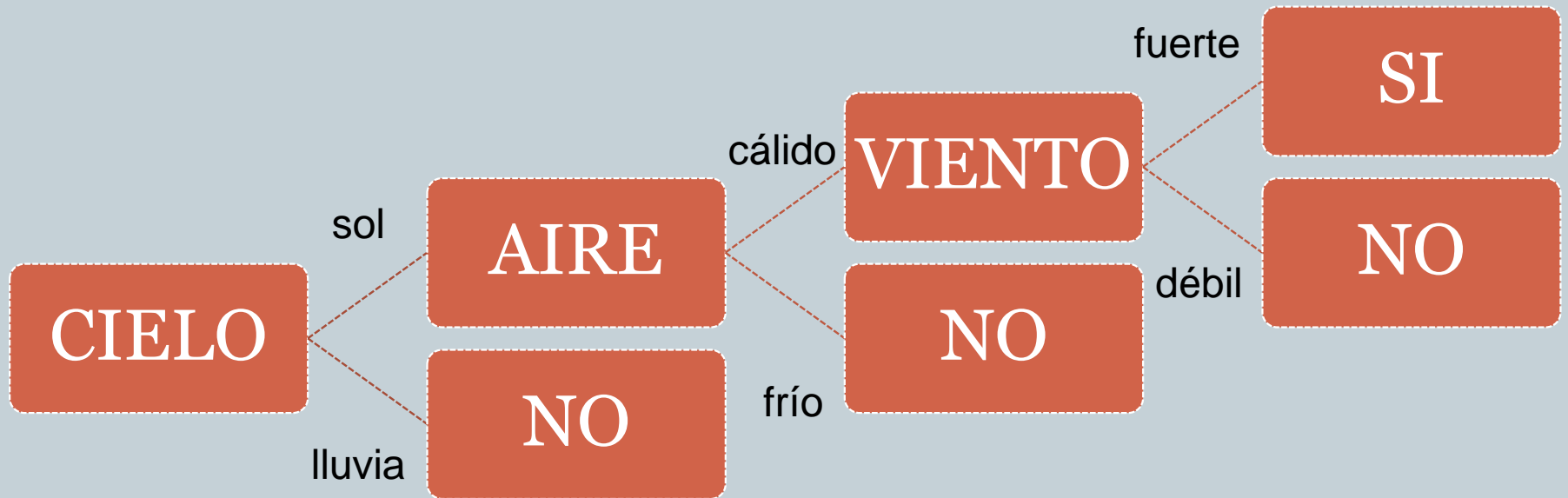
Aprendizaje de conceptos

5

- Métodos de búsqueda de las hipótesis verdaderas:
 - FIND-S:
Usando **solo** los ejemplos de la muestra de entrenamiento encuentra **una** hipótesis verdadera, la hipótesis específica maximal
 $\{ \text{sol, cálido, ?, fuerte, ?, ?} \}$
 - Listado y eliminación:
Barre el árbol completo de hipótesis comenzando desde la más general para encontrar **todas** las hipótesis verdaderas
 - Eliminación de candidatos:
Se realiza la búsqueda exhaustiva en una versión acotada del espacio de hipótesis

Árbol asociado

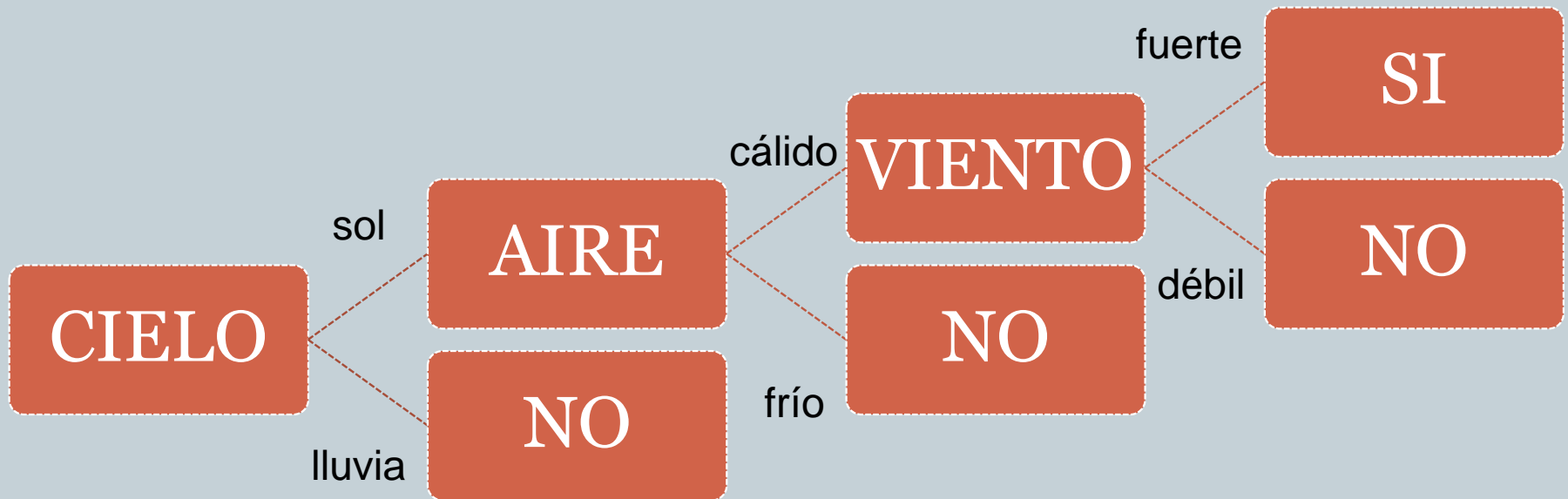
6



Deporte: tenis

7

- Si en vez de ser un deporte de vela fuese tenis
- Con los mismos datos, como quedaría el árbol?



Aprendizaje de conceptos

8

- Implementación del clasificador obtenido con los algoritmos:
 - Memorizador (Rote Learning):
Responde sólo si la nueva instancia está previamente almacenada.
No tiene sesgo inductivo.
 - Eliminación de candidatos:
Responde sólo si para la nueva instancia hay acuerdo entre todas las hipótesis. Hay sesgo inductivo: supone que el espacio versionado contiene el concepto
 - FIND-S:
Responde siempre ante una nueva instancia. Árbol de decisión.
El sesgo inductivo es muy grande: La hipótesis mínima maximal es el concepto
- El aprendizaje de conceptos adolece de intolerancia al ruido !

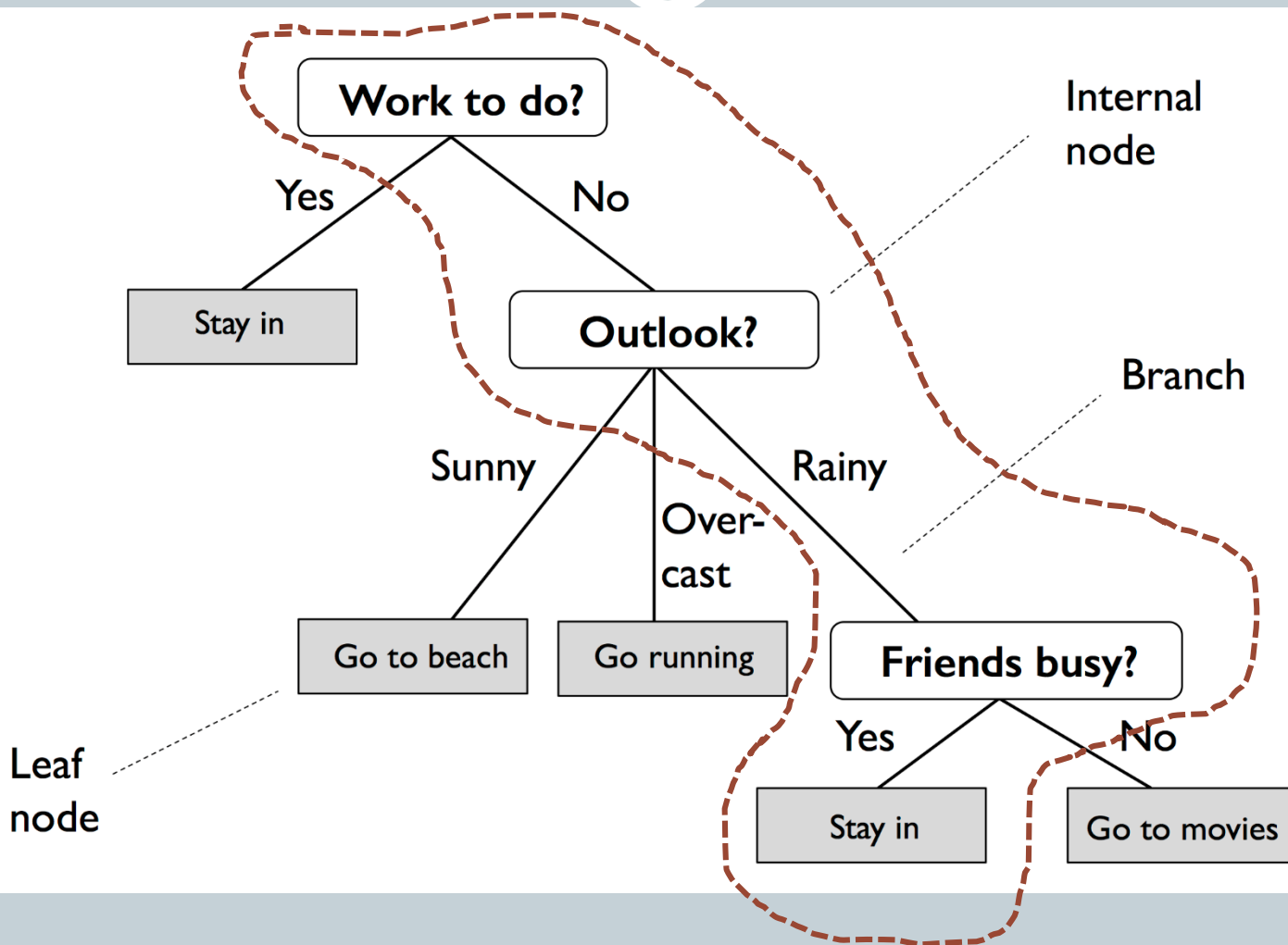
Arboles

9

- Árboles de decisión son reglas anidadas “if-else”.
- Una regla es una intersección de condiciones
 - Rule 1 = (if $x = 1$) \cap (if $y = 2$) \cap ...
- Reglas múltiples pueden unirse para formar un conjunto de reglas que pueden ser usadas para otorgar etiquetas.
 - Class 1 = if (Rule 1=True) \cup (Rule 2=True) \cup ...
- Ejemplo
 - Cada nodo hoja en la figura siguiente representa un conjunto de reglas . La figura total es el clasificador
 - (Work to do? = False) \cap (Outlook? = Rainy) \cap (Friends busy? = Yes)

Árbol categórico

10



Práctico segunda semana A

11

En un consultorio oftalmológico la secretaria entrega a cada paciente la siguiente ficha, antes ser atendido, para que sea completada por el profesional:

paciente	(nombre.apellido)
edad	(años)
prescripción	(miopía / hipermetropía)
astigmatismo	(sí / no)
lagrimeo	(normal/reducido)

En cada turno de trabajo se atienden seis pacientes y a continuación se presentan las fichas completadas por el oculista, adecuadamente anonimizadas:

En cada turno de trabajo se atienden seis pacientes y a continuación se presentan las fichas completadas por el oculista, adecuadamente anonimizadas:

Primer turno

paciente	01	02	03	04	05	06
edad	22	71	41	32	68	72
prescripción	miopía	miopía	hipermetropía	miopía	hipermetropía	miopía
astigmatismo	no	sí	no	sí	sí	no
lagrimeo	reducido	normal	reducido	normal	reducido	normal

Segundo turno

paciente	07	08	09	10	11	12
edad	62	42	38	43	28	37
prescripción	hipermetropía	miopía	miopía	miopía	hipermetropía	miopía
astigmatismo	sí	sí	no	no	sí	sí
lagrimeo	normal	reducido	reducido	normal	normal	normal

Tercer turno

paciente	13	14	15	16	17	18
edad	45	37	68	18	51	15
prescripción	hipermetropía	hipermetropía	hipermetropía	miopía	miopía	hipermetropía
astigmatismo	sí	no	no	sí	no	no
lagrimeo	normal	normal	normal	reducido	reducido	normal

Cuarto turno

paciente	19	20	21	22	23	24
edad	63	30	74	44	15	17
prescripción	miopía	miopía	hipermetropía	hipermetropía	hipermetropía	hipermetropía
astigmatismo	sí	no	no	sí	no	sí
lagrimeo	reducido	normal	reducido	reducido	reducido	reducido

Una vez completada la ficha, los pacientes se dirigen luego al consultorio del contactólogo para que les recomienden o no el uso de lentes de contacto. Al final del día el resumen de recomendaciones se recopila en la siguiente tabla:

paciente	recomendación	paciente	recomendación	paciente	recomendación
01	nada	09	nada	17	nada
02	duras	10	blandas	18	blandas
03	nada	11	nada	19	nada
04	duras	12	duras	20	blandas
05	nada	13	nada	21	nada
06	blandas	14	blandas	22	nada
07	nada	15	blandas	23	nada
08	nada	16	nada	24	nada