

Resumen-tema-4.pdf



Carmenroman5



Sistemas Inteligentes



3º Grado en Ingeniería de la Salud



Escuela Técnica Superior de Ingeniería Informática
Universidad de Sevilla

RALPH FIENNES ANYA TAYLOR-JOY NICHOLAS HOULT

el **MENÚ**

YA EN CINES

COMPRAR ENTRADAS

©2022 20th Century Studios.



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

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



Dí adiós a la publi
en los apuntes y
en la web



Participa gratis
en todos los
sorteos



Descarga
carpetas
completas

estudia sin publi
WUOLAH PRO

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

TEMA 4 : APRENDIZAJE SUPERVISADO E INDUCTIVO

DEFINICIÓN

Cualquier cambio en un sistema que le permita realizar la misma tarea de manera más eficiente la próxima vez

→ **APRENDIZAJE AUTOMÁTICO** : construye programas que mejoran automáticamente con la experiencia

→ **TIPOS DE APRENDIZAJE**

1. **SUPERVISADO**: conjunto de datos de los que conocemos como se comportan respecto a lo que se quiere modelar. Se pretende predecir el comportamiento sobre instancias nuevas
2. **NO SUPERVISADO**
3. **CON ESFUERZO**

→ **TIPOS DE PARADIGMAS**

1. Aprendizaje por memorización
2. Aprendizaje inductivo
3. Clasificación
4. Aprendizaje por analogía
5. Descubrimiento
6. Redes neuronales

CLASIFICACIÓN

1. Tenemos clases. $C = \{c_1, \dots, c_K\}$
2. Se trata de aprender un modelo que nos permita asignarle una clase a nuevos datos
3. Conjunto de entrenamiento :

$$D = \{(x^j), (c^j)\} \Psi, j=1 \dots N$$

→ Clasificación binaria : $C = \{0, 1\} \Psi, \{+, -\} \Psi$

4. Buscar una función $h: A^n \rightarrow C$

↓
Ajuste el conjunto de entrenamiento y sirva para clasificar nuevas instancias.

RALPH
FIENNES ANYA
TAYLOR-JOY NICHOLAS
HOULT

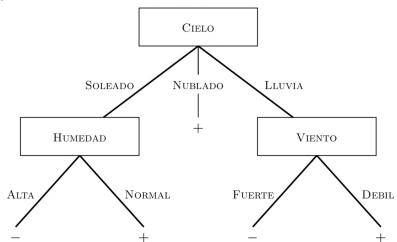
el **MENÚ**

Escrupulosamente preparado. Brillante ejecución.



ÁRBOLES DE DECISIÓN

- Nodos interiores: atributos
- Arcos: posibles valores nodo origen
- Hojas: valor de la clasificación
- Representación de una función objetivo



Ejemplo Árbol Decision

→ Disyunción reglas proposicionales

$$\begin{aligned} & (\text{CIELO}=\text{SOLEADO} \wedge \text{HUMEDAD}=\text{ALTA} \rightarrow \text{JUGARTENIS}= -) \\ \vee & (\text{CIELO}=\text{SOLEADO} \wedge \text{HUMEDAD}=\text{NORMAL} \rightarrow \text{JUGARTENIS}= +) \\ \vee & (\text{CIELO}=\text{NUBLADO} \rightarrow \text{JUGARTENIS}= +) \\ \vee & (\text{CIELO}=\text{LLUVIOSO} \wedge \text{VENTO}=\text{FUERTE} \rightarrow \text{JUGARTENIS}= -) \\ \vee & (\text{CIELO}=\text{LLUVIOSO} \wedge \text{VENTO}=\text{DEBIL} \rightarrow \text{JUGARTENIS}= +) \end{aligned}$$

→ ¿Cómo saber que atributo clasifica mejor?

1. Entropía de un conjunto de ejemplos D

$$\text{Ent}(D) = - \frac{|P|}{|D|} \cdot \log_2 \frac{|P|}{|D|} - \frac{|N|}{|D|} \cdot \log_2 \frac{|N|}{|D|}$$

- Subconjunto de ejemplos positivos y negativos de D
Notación:

$$\begin{array}{c} \text{Ent}([p^+, n^-]) \\ \downarrow \quad \downarrow \\ [P] \quad [N] \end{array}$$

2. Intuición: mide la ausencia de homogeneidad de la clasificación.

→ Teoría de la información: cantidad media de información (bits) necesaria para codificar la clasificación de un ejemplo.

→ Ganancia de información

- Nodos con menos entropía (árboles pequeños)
- Entropía esperada tras usar un atributo A en el árbol.

$$\sum \frac{|D_v|}{|D|} \cdot \text{Ent}(D_v)$$

Subconjuntos de ejemplos de D con valor del atributo A igual a v

- Ganancia de información esperada tras usar un atributo A

$$\text{Ganancia}(D, A) = \text{Ent}(D) - \sum \frac{|D_v|}{|D|} \cdot \text{Ent}(D_v)$$

BÚSQUEDA Y SESGO INDUCTIVO

→ Búsqueda en un espacio de hipótesis

- Espacio de todos los árboles de decisión
- Un único árbol candidato en cada paso
- Sin retroceso
- Búsqueda en escalada
- Decisiones tomadas a partir de conjuntos de ejemplos

→ Sesgo inductivo

- Árboles cortos sobre largos
- Sesgo preferencial
- Principio navaja Occam: la solución más sencilla, la correcta



¿Quién te conoce mejor?

Tu madre

Tus amigos

Tu Wrapped

Si dudas, echa un ojo a tu Wrapped

#SPOTIFYWRAPPED



MEDIDA DEL RENDIMIENTO DEL APRENDIZAJE

1) CONJUNTO DE ENTRENAMIENTO Y PRUEBA:

- Aprender con el conjunto de entrenamiento
- **Medida de rendimiento:** proporcion de ejemplos bien clasificados en el conjunto de prueba

2) REPETICION DE ESTE PROCESO

- Curva de aprendizaje
- **Estratificación:** cada clase correctamente representada en el entrenamiento y en la prueba

3) VALIDACION CRUZADA

- Dividir en K partes y hacer K aprendizajes. Finalmente hacer la media de los rendimientos

SOBREALJUSTE Y RUIDO

→ Actua mejor sobre la distribucion completa de instancias

→ **RUIDOS:** ejemplos incorrectamente clasificados → **Causa sobreajuste**

↓ OTRAS

- Atributos no relevantes
- Conjuntos entrenamiento pequeños

→ MANERAS EVITAR SOBREALJUSTE

1. Pasar el desarrollo del arbol antes de que se ajuste perfectamente
2. Podar el arbol despues

Quééééédate
y descubre
tu resumen
del año.



#SPOTIFYWRAPPED

ATRIBUTOS CON VALORES CONTINUOS

→ Reemplazamos los atributos continuos por booleanos, introduciendo un valor C

· A continuo



Temperatura	50	52	60	68	70	78	84
Clase	+	+	-	-	-	+	+

- ▶ Los candidatos a umbral C son los valores adyacentes con distinta clase
 - ▶ $56 = (52 + 60)/2$ y $74 = (70 + 78)/2$.
- ▶ Seleccionamos el umbral con máxima ganancia
- ▶ El nuevo atributo $A_{<C}$ compite con los restantes.
- ▶ El proceso se realiza a cada paso eligiendo el mejor umbral en el conjunto de entrenamiento.

ALGORITMO DE COBERTURA

→ Diferentes criterios para elegir la mejor condición en cada vuelta del bucle interno:

- Se añade la condición con mayor frecuencia relativa
- Se añade la de mayor ganancia de información

$$P \cdot \left(\log_2 \frac{P'}{t'} - \log_2 \frac{P}{t} \right)$$

frecuencia relativa antes
de la condición

↓ frecuencia relativa después
de añadir la condición

→ Las reglas se ajustan perfectamente al conjunto de información (peligro sobreajuste)

VISIÓN ESTADÍSTICA DEL PORCENTAJE

→ El problema se plantea con:

1. Conjunto datos d : evidencias observadas

2. Hipótesis H : modelos probabilísticos de como funciona el dominio.

→ Las decisiones se tomaran calculando probabilidades condicionales → Se busca la hipótesis mas probable , dada las observaciones d

FORMLULACIÓN REGLA DE BAYES

$$\rightarrow P(b|a) = \frac{P(a|b) \cdot P(b)}{P(a)}$$

→ Versión con normalización: $P(Y|X) = \alpha \cdot P(X|Y) \cdot P(Y)$

→ Generalización con un conjunto de e observaciones:

$$P(Y|X, e) = \alpha \cdot P(X|Y, e) P(Y, e)$$

→ Marco general

- La información de entrada es:
 - Datos observados, d
 - Probabilidades a priori , $P(h)$
 - Verosimilitud, $P(d|h)$

· Se calcula $P(h|d)$

$$P(h|d) = \alpha \cdot P(d|h) \cdot P(h)$$

→ Relación causa - efecto

$$P(\text{causa} | \text{Efecto}) = \alpha \cdot P(\text{Efecto} | \text{Causa}) \cdot P(\text{Efecto})$$

→ Clasificadores Bayesianos

· VMAP : valor de clasificación asignado a una nueva instancia
argmax $P(v_j | a_1, \dots, a_n)$

↓ Aplicando el teorema de Bayes

$$V_{MAP} = \text{argmax} P(a_1, \dots, a_n | v_j) \cdot P(v_j)$$

→ Despues estimar las probabilidades → PROBLEMA: son muchas para hacerlo bien

→ Clasificadores Naive-Bayes

- Simplificamos suponiendo que los atributos son independientes del valor de clasificación
- Valor clasificación: $VNB = \operatorname{argmax} P(V_j) \prod_i P(a_i | V_j)$

- Solo tenemos que estimar $P(V_j)$ y $P(a_i | V_j)$

$$P(V_j) = \frac{\#(V=V_j)}{N}, \quad P(a_i | V_j) = \frac{\#(A_i=a_i, V=V_j)}{\#(V=V_j)}$$

\hookrightarrow N° total ejemplos \hookrightarrow N° ejemplos clasificados

SOLUCIONES A LOS RIESGOS DE ESTIMACIONES BAJAS

I. SUAVIZADO LAPLACE

$$\frac{n + k}{n + K |A_i|}$$

\hookrightarrow n° de valores distintos que puede tomar A_i

CLASIFICACIÓN MEDIANTE VECINO MÁS CERCANO

- Obtiene la clasificación de un nuevo ejemplo a partir de las categorías de los ejemplos más cercanos.
- Manejo de distancias
- Mayoria casos distancia euclídea



¿Quién te conoce mejor?

Tu madre Tus amigos Tu Wrapped

#SPOTIFYWRAPPED

Si dudas, echa un ojo a tu Wrapped

¿Tu madre,
tus amigos
o tu
Wrapped?

Descubre
quién te conoce
mejor en tu
resumen del año.



MÁS INFORMACIÓN



ALGORITMO K-NN

1. KNN

- No paramétrico
- Perezoso
- Fase de entrenamiento rápida, pero necesarios todos los datos para hacer la predicción

2. VECINO MÁS PROXIMO

- Consideramos un solo vecino $K=1$
- Fase entrenamiento: guardamos el conjunto de entrenamiento y las etiquetas asociadas a cada muestra
- Fase predicción: (nuevas muestras): dada muestra x , buscamos la más cercana x_i y asociamos a x la etiqueta y_i

3. K VECINOS MÁS PROXIMOS

- $K > 1$
- Fase entrenamiento: guardamos el conjunto de entrenamiento y las etiquetas asociadas a cada muestra
- Fase predicción: (nuevas muestras): dada muestra x , buscamos en el conjunto las K más cercanas. Nos quedamos con la que tengan más votos.

MEDIDAS DE PROXIMIDAD

→ DISTANCIA: específico de cada problema. La más usada es la euclídea

Características

Tipos

- | | |
|---|--|
| <ul style="list-style-type: none"> ► (SIM) Simetría: $d(x, y) = d(y, x) \quad \forall x, y$ ► (DNN) Definida no negativa: $d(x, y) \geq 0 \quad \forall x, y$ ► (DTR) Desigualdad triangular: $d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z$ ► (REF) Reflexividad: $d(x, x) = 0 \quad \forall x$ ► (IND) Indistinguibilidad: $d(x, y) = 0 \Rightarrow x = y \quad \forall x, y$ ► Otras... | <ul style="list-style-type: none"> ► Premétrica: DNN + REF ► Semimétrica: DNN + REF + IND + SIM ► Pseudométrica: DNN + REF + SIM + DTR ► Quasimétrica: DNN + REF + IND + DTR ► Pseudoquasimétrica: DNN + REF + DTR ► Métrica: DNN + REF + IND + SIM + DTR ► Intuición para métricas débiles: Energía consumida para ir de un punto a otro |
|---|--|

→ MÁS USADAS

- Euclídea: $d_e(x, y) = \sqrt{(x_1 - y_1)^2}$
- Manhattan: $d_m(x, y) = |x_1 - y_1|$
- Hamming: nº componentes en los que se difiere

BOLETÍN TEMA 4

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

¿No te llega para pagar Wuolah Pro? ¿Un año sin anuncios gratis?

Reservados todos los derechos.
No se permite la explotación económica ni la transformación de esta obra. Queda permitida la impresión en su totalidad.

WUOLAH

Llévate 1 año de WUOLAH PRO con BBVA. ¿Cómo? ¡+Info aquí!