MACHINE LEARNING

CIENCIAS DE LA COMPUTACIÓN
M.Sc. Richard F. Fernández Vásquez

¿Qué es una probabilidad?

Logro Unidad 2

Al finalizar la unidad, el alumno entiende e implementa soluciones de Machine Learning basados en algoritmos tradicionales de aprendizaje supervisado, interpreta los resultados obtenidos y obtiene conclusiones coherentes y adecuadas.

Contenido 4

- Probabilidades
- 1 Regla
- Clasificador Naive Bayes
- Caso de Aplicación

Contenido 4

- Probabilidades
- 1 Regla
- Clasificador Naive Bayes
- Caso de Aplicación

Experimento aleatorio

Cumple con las siguientes características:

- Se puede repetir indefinidamente donde los resultados dependen del azar, por lo que no se pueden predecir con certeza.

- Se puede describir el conjunto de todos los resultados posibles.

- Cuando se repite en un gran número de veces, aparece un modelo definido de regularidad.

- Se le suele simbolizar como Ɛ.

Espacio muestral

- Es el conjunto de todos los posibles resultados de un experimento aleatorio.

- Se le suele simbolizar como S o Ω.

Evento

- Un evento es un subconjunto del espacio muestral.
- Ejemplo:

E= Al lanzar un dado una vez, que en la cara superior aparezca un número par.

S = {1, 2, 3, 4, 5, 6}

E = {2, 4, 6}

P(E) = n(E)/n(S) = 3/6

Axiomas de probabilidad

Sea P una función que asocia a cada evento E de S un número real:

1. Para algún evento E, 0 <= P(E) <= 1
2. P(S) = 1
3. P(A ∪ B) = P(A) + P(B), si A ∩ B = ϕ

En general: para alguna secuencia de eventos mutuamente excluyentes En, n>=1, tal que Ei ∩ Ej=∅ → $P(\bigcup_{n=1}^{\infty} E_n)=\sum_{n=1}^{\infty} P(E_n)$

S = {1, 2, 3, 4, 5, 6}

A = {2, 4, 6}

B = {1, 3, 5}

P(A ∩ B) = 0 => A y B son eventos disjuntos o mutuamente excluyentes

Propiedades

1. Teorema: Probabilidad de un evento nulo: $P(\emptyset)=0$
2. Probabilidad del evento complemento: $P(E')=1-P(E)$
3. Probabilidad de eventos incluidos: **Si $A\subset B$** entonces $P(A)\leq P(B)$
4. La probabilidad de un evento está entre cero y uno: $0\leq P(E) \leq 1$
5. Teorema: Regla aditiva de probabilidad de eventos:

$$P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

Propiedades

1. Teorema: Probabilidad de un evento nulo: P(Ø)=0

S = S U ϕ

<u>P(S)</u> = P(S U ϕ) = <u>P(S) + P(ϕ)</u> => S y ϕ son disjuntos o m.e.

1 = 1 + P(ϕ)

P(ϕ) = 0

Propiedades

2. Probabilidad del evento complemento: P(E')=1-P(E)

S = E U E'

P(S) = P(E U E') => E y E' son disjuntos o m.e.

P(S) = P(E) + P(E')

1 = P(E) + P(E')

P(E') = 1 – P(E)

Propiedades

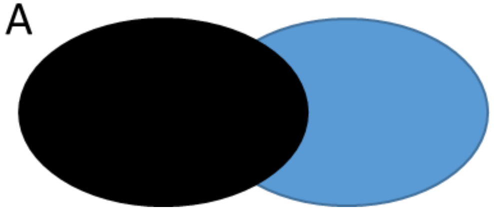
4. La probabilidad de un evento está entre cero y uno: $0 \leq P(E) \leq 1$

$\Phi \subset E \subset S$

$P(\Phi) \leq P(E) \leq P(S)$

$0 \leq P(E) \leq 1$

Propiedades



5. Teorema: Regla aditiva de probabilidad de eventos:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$A \cup B = A \cup (B - A)$

$P(A \cup B) = P(A \cup (B - A)) = P(A) + P(B\text{-}A) \Rightarrow P(A \cup B) = P(A) + P(B\text{-}A)$ ... (1)

$B = (A \cap B) \cup (B - A)$

$P(B) = P(A \cap B) + P(B - A) \Rightarrow P(B) - P(A \cap B) = P(B - A)$ ... (2)

De (1) en (2)

$P(A \cup B) - P(A) = P(B) - P(A \cap B) \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Propiedades

6. $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$

7. Teorema: Sea S un espacio muestral y A un evento de S, $A \subset S$

Entonces: $P(A) = P(A1)+P(A2)+\ldots+P(Ak)$,

$P(A) = \sum_{i=1}^{k} P(A_i)$, donde Ai son eventos disjuntos cuya unión es A

Propiedades

7. Teorema: Sea S un espacio muestral y A un evento de S, A⊂S

Entonces: P(A) = P(A1)+P(A2)+…+P(Ak),

P(A) = $\sum_{i=1}^{k} P(A_i)$, donde Ai son eventos disjuntos cuya unión es A

A = A1 U A2 U A3 U … U Ak

P(A) = P(A1) + P(A2) + P(A3) + … + P(Ak)

P(A) = $\sum_{i=1}^{k} P(A_i)$

Probabilidad condicional

La probabilidad condicional se refiere a hallar la probabilidad de un evento conociendo cierta información (condición) y se define como:

$$P\left(A/B\right) = \frac{P(A \cap B)}{P(B)}$$

Representa la probabilidad condicional de A, dado que ocurrió el evento B.

Donde $P(B) > 0$.

Axiomas de probabilidad

Sea **S**: Espacio muestral

A, B: Eventos de S

1. $P(S|B) = 1$
2. $P(A \cup C \mid B) = P(A|B) + P(C|B)$

Axiomas de probabilidad

Sea **S**: Espacio muestral

A, B: Eventos de S

1. $P(S|B) = 1$

$$P(S|B) = P(S \cap B)/P(B) = P(B) / P(B) = 1$$

$$P(S|B) = 1$$

Axiomas de probabilidad

Sea **S**: Espacio muestral

A, B: Eventos de S

2. $P(A \cup C \mid B) = P(A|B) + P(C|B)$

$$P(A \cup C \mid B) = P((A \cup C) \cap B)/P(B) = P((A \cap B) \cup (C \cap B))/P(B)$$

$$= P(A \cap B)/P(B) \cup (C \cap B)/P(B)$$

$$= P(A|B) + P(C|B)$$

Propiedades

1. $0 \leq P(B|A) \leq 1$
2. $P(S|A) = 1$
3. $P(A \cup B|C) = P(A|C) + P(B|C)$, si A y B son disjuntos o m.e.

Probabilidad condicional

La mayoría de las estaciones de servicio venden tres tipos de gasolina: 84 octanos, 95 octanos y 97 octanos. Con frecuencia, alguna de cada está enriquecida con un aditivo. La tabla siguiente ilustra los porcentajes de clientes que prefieren cada tipo.

	90 octanos (B)	95 octanos (C)	97 octanos (D)	Total
Con Aditivo (A)	0.05	0.10	0.05	0.20
Sin Aditivo (A’)	0.15	0.40	0.25	0.80
Total	0.20	0.50	0.30	1.00

a) ¿Cuál es la probabilidad de que haya comprado gasolina con aditivo o no sea de 95 octanos?

b) Si el cliente no compró gasolina de 95 octanos, ¿cuál es la probabilidad de que hay comprado gasolina de 97 octanos?

c) Si el cliente no compró gasolina de 90 0ctanos, ¿cuál es la probabilidad de que haya comprado gasolina sin aditivo?

	90 octanos (B)	95 octanos (C)	97 octanos (D)	Total
Con aditivo(A)	0,05	0,10	0,05	**0,20**
Sin aditivo (A')	0,15	0,40	0,25	**0,80**
Total	**0,20**	**0,50**	**0,30**	**1,00**

a) ¿Cuál es la probabilidad de que haya comprado gasolina con aditivo o no sea de 95 octanos?

$$P(A \cup C') = P(A) + P(C') - P(A \cap C') = 0{,}20 + 0{,}50 - (0{,}05 + 0{,}05) = 0{,}60$$

b) Si el cliente no compró gasolina de 95 octanos, ¿cuál es la probabilidad de que hay comprado gasolina de 97 octanos?

$$P\left(D/_{C'}\right) = \frac{P(D \cap C')}{P(D')} = \frac{0{,}30}{0{,}50} = 0{,}60$$

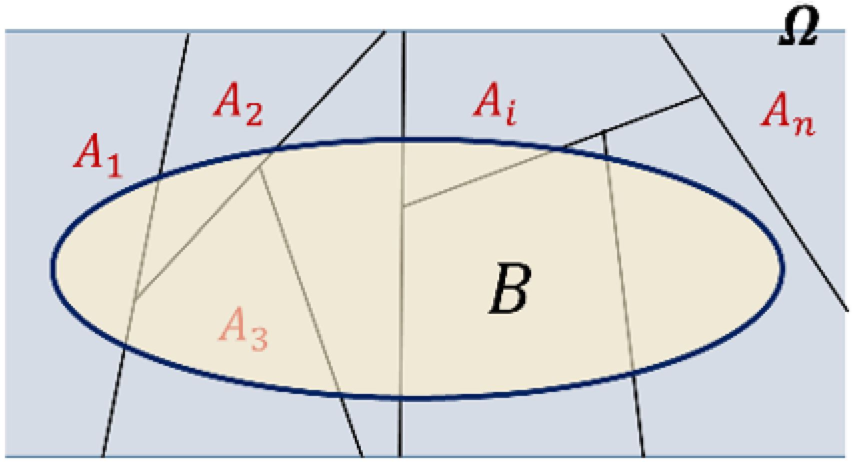
c) Si el cliente no compró gasolina de 90 0ctanos, ¿cuál es la probabilidad de que haya comprado gasolina sin aditivo?

$$P\left(A'/_{B'}\right) = \frac{P(A' \cap B')}{P(B')} = \frac{0{,}65}{0{,}80} = 0{,}8125$$

Probabilidad Total

Teorema: (Probabilidad total) Suponga que los eventos $A_1, A_2, \ldots, A_k$ forman una partición de Ω, es decir, $A_1 \cup A_2 \cup \ldots \cup A_k = \Omega$, $A_i \neq \emptyset$ y $A_i \cap A_j = \emptyset \quad \forall i \neq j$. Entonces para cualquier evento $E \subset \Omega$ se tiene:

$$P(E) = \sum_{i=1}^{k} P(A_i) \cdot P(E/A_i)$$



Teorema de Bayes

Si $A_1, A_2, \ldots, A_k$ es una partición de Ω, es decir, $A_1 \cup A_2 \cup \ldots \cup A_k = \Omega$, $A_i \neq \emptyset$ y $A_i \cap A_j = \emptyset$ $\forall\, i \neq j$. Entonces para cualquier evento $B \subset \Omega$ se tiene:

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

$$P(A_i/B) = \frac{P(B \cap A_i)}{P(B \cap A_1) + P(B \cap A_2) + \ldots + P(B \cap A_k)}$$

$$P(A_i/B) = \frac{P(B/A_i) \cdot P(A_i)}{P(B/A_1) \cdot P(A_1) + P(B/A_2) \cdot P(A_2) + \ldots + P(B/A_k) \cdot P(A_k)}$$

Teorema de Bayes

Para dos sucesos A y B,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$p(A|B)$: Probabilidad a posteriori
$p(B|A)$: Función de verosimilitud
$p(A)$: Probabilidad a priori

Contenido 4

- Probabilidades
- 1 Regla
- Clasificador Naive Bayes
- Caso de Aplicación

1 Regla (1R)

- 1R: es equivalente a un árbol de decisión de 1nivel
    - Todas las reglas usan solamente 1 atributo
    - Atributo debe ser (transformado) de tipo categórico
- Versión básico
    - Una rama para cada valor del atributo
    - Para cada rama, colocar la clase más frecuente
    - Para cada rama, calcular la tasa de error de clasificación
    - Proporción de ejemplos que no pertenecen a la clase más frecuente
    - Escoger el atributo con la menor tasa de error de clasificación

1 Regla (1R)

- Algoritmo 1R:

    Para cada valor del atributo generar una regla de la siguiente manera:

        Contar la frecuencia de cada clase

        Encontrar la clase más frecuente

        Formar una regla que le asigna la clase más frecuente a dicho atributo-valor

- Escoger las reglas con la menor tasa de error de clasificación

1 Regla (1R)

Outlook	Temp	Humidity	Windy	Play
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Cool	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

Attribute	Rules	Errors	Total Errors
Outlook	Sunny → No	2/5	4/14
	Overcast → Yes	0/4	
	Rainy → Yes	2/5	
Temp	Hot → No*	2/4	5/14
	Mild → Yes	2/6	
	Cool → Yes	1/4	
Humidity	High → No	3/7	4/14
	Normal → Yes	1/7	
Windy	False → Yes	2/8	5/14
	True → No*	3/6	

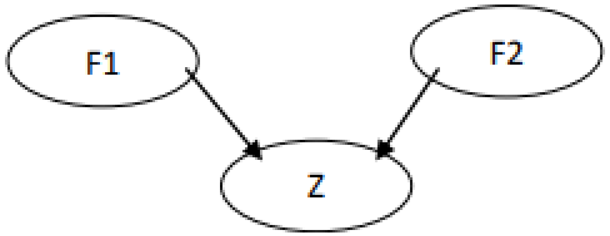
Contenido 4

- Probabilidades
- 1 Regla
- Clasificador Naive Bayes
- Caso de Aplicación

Redes Bayesianas

Una red bayesiana es un grafo dirigido en el que cada nodo contiene una información probabilística.

Consideremos una variable aleatoria Z dependiente de otras dos F1 y F2:



Clasificadores Naive Bayes

La regla de Bayes establece que, si tenemos una hipótesis "H" sustentada para una evidencia "E", entonces:

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)}$$

$p(H|E)$: Probabilidad a posteriori

$p(E|H)$: Función de verosimilitud

$p(H)$: Probabilidad a priori

Clasificadores Naive Bayes

Ejemplo: Una compañía de seguros dispone de los siguientes datos sobre sus clientes, clasificados en buenos y malos clientes:

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

- La hipótesis "H" es que "buen cliente" sea Si o No.

- La evidencia E es una combinación de los valores de los atributos: edad, hijos, practica deporte y salario del dato nuevo, por lo que su probabilidad se obtiene multiplicando las probabilidades de estos valores:

$$p(si|E) = \frac{[p(edad|si).p(hijos|si).p(practica\ deporte|si).p(salario|si)].p(si)}{p(E)}$$

$$p(no|E) = \frac{[p(edad|no).p(hijos|no).p(practica\ deporte|no).p(salario|no)].p(no)}{p(E)}$$

Clasificadores Naive Bayes

Tenemos un nuevo caso con las siguientes características:

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
11	Mayor	No	No	Medio	?

$$p(si|E) = [p(edad|si).p(hijos|si).p(practica\ deporte|si).p(salario|si)].p(si)$$

$$p(no|E) = [p(edad|no).p(hijos|no).p(practica\ deporte|no).p(salario|no)].p(no)$$

Clasificadores Naive Bayes

$p(buen\ cliente = si)$ = 7 / 10

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(edad = mayor|buen\ cliente = si)$ = 3 / 7

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(hijos = no|buen\ cliente = si)$ = 2 / 7

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(practica\ deporte = no|buen\ cliente = si)$ = 4 / 7

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(salario = medio|buen\ cliente = si)$ = 3 / 7

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(buen\ cliente = no)$ = 3 / 10

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(edad = mayor|buen\ cliente = no)$ = 1 / 3

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(hijos = no|buen\ cliente = no)$ = 2 / 3

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(practica\ deporte = no|buen\ cliente = no)$ = 2 / 3

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

$p(salario = medio|buen\ cliente = no)$ = 2 / 3

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
1	Joven	Si	No	Alto	Si
2	Joven	No	No	Medio	No
3	Joven	Si	Si	Medio	No
4	Joven	Si	No	Bajo	Si
5	Mayor	Si	No	Bajo	Si
6	Mayor	No	Si	Medio	Si
7	Joven	No	Si	Medio	Si
8	Joven	Si	Si	Alto	Si
9	Mayor	Si	No	Medio	Si
10	Mayor	No	No	Bajo	No

Clasificadores Naive Bayes

Tenemos un nuevo caso con las siguientes características:

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
11	Mayor	No	No	Medio	?

Clasificadores Naive Bayes

- Las probabilidades se pueden determinar de la siguiente manera:

$p(si)$ = 7 / 10

$p(edad = mayor|si)$ = 3 / 7

$p(hijos = no|si)$ = 2 / 7

$p(practica\ deporte = no|si)$ = 4 / 7

$p(salario = medio|si)$ = 3 / 7

$p(no)$ = 3 / 10

$p(edad = mayor|no)$ = 1 / 3

$p(hijos = no|no)$ = 2 / 3

$p(practica\ deporte = no|no)$ = 2 / 3

$p(salario = medio|no)$ = 2 / 3

$p(si|E) = 0.0210$

$p(no|E) = 0.0296$

Clasificadores Naive Bayes

Tenemos un nuevo caso con las siguientes características:

ID	Edad	Hijos	Practica Deporte	Salario	Buen Cliente
12	Joven	Si	Si	Alto	No

Clasificadores Naive Bayes

Naive Bayes tienes 3 maneras de ser usado:

- Gaussian: asume que los atributos tienen una distribución normal.
- Multinomial: usado para atributos contables discretos. Por ejemplo: frecuencia de aparición de palabras en texto.
- Bernoulli: usado para atributos binarios.

CIERRE

Naive Bayes

¿Qué es una probabilidad?

Teorema de Bayes

CONSULTAS

pcsirife@upc.edu.pe