09481: Inteligencia Artificial

Profesor del curso: Breyner Posso, Ing. M.Sc.
e-mail: breyner.posso1@u.icesi.edu.co

Programa de Ingeniería de Sistemas.
Departamento TIC.
Facultad de Ingeniería.
Universidad Icesi.
Cali, Colombia.

Agenda

1. Introducción
2. Repaso de probabilidad.
3. Clasificador Bayes Ingenuo (Naïve Bayes).

1. Introducción

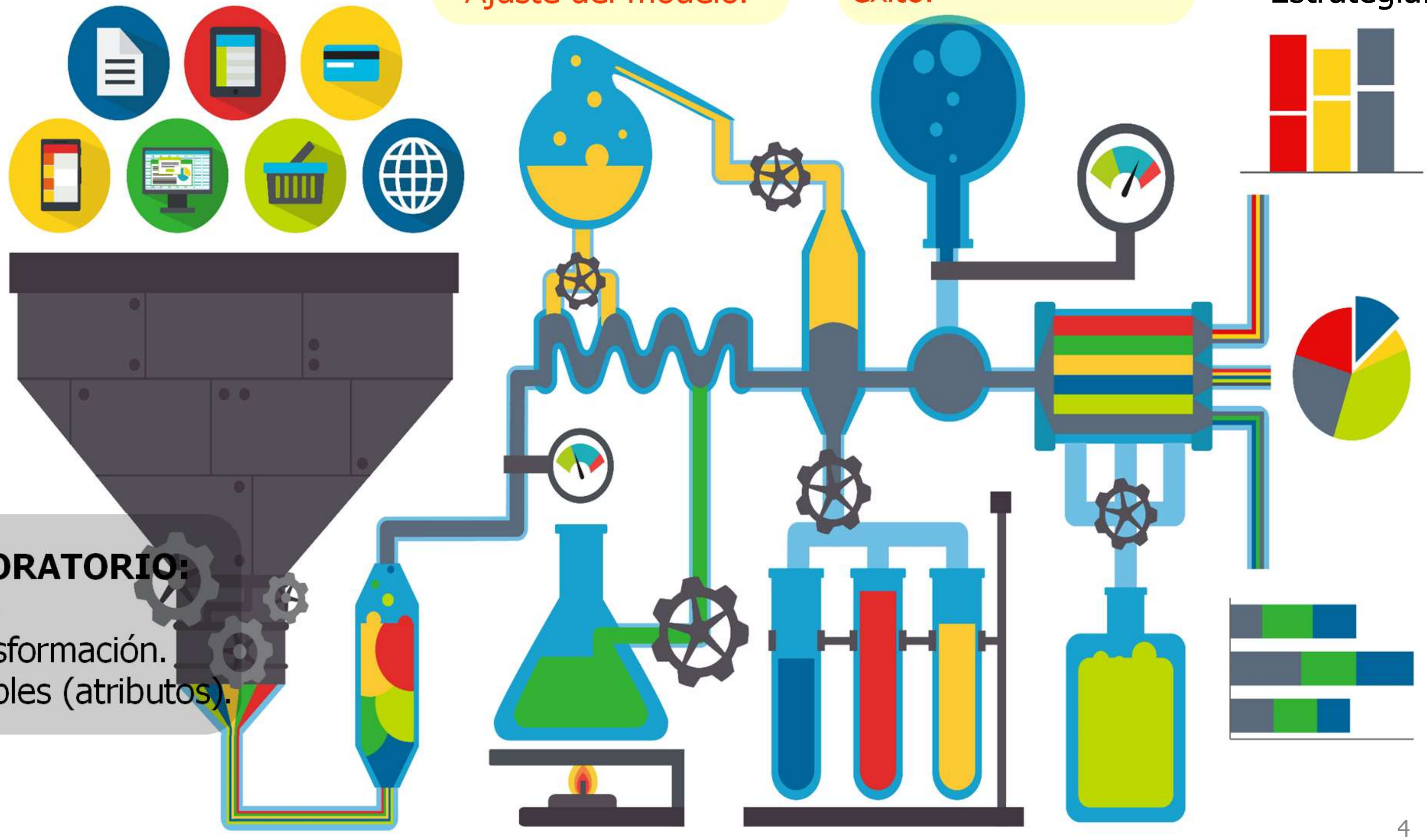
DATOS:
Materia prima.

MODELO:
Implementación del modelo de analítica.
Ajuste del modelo.

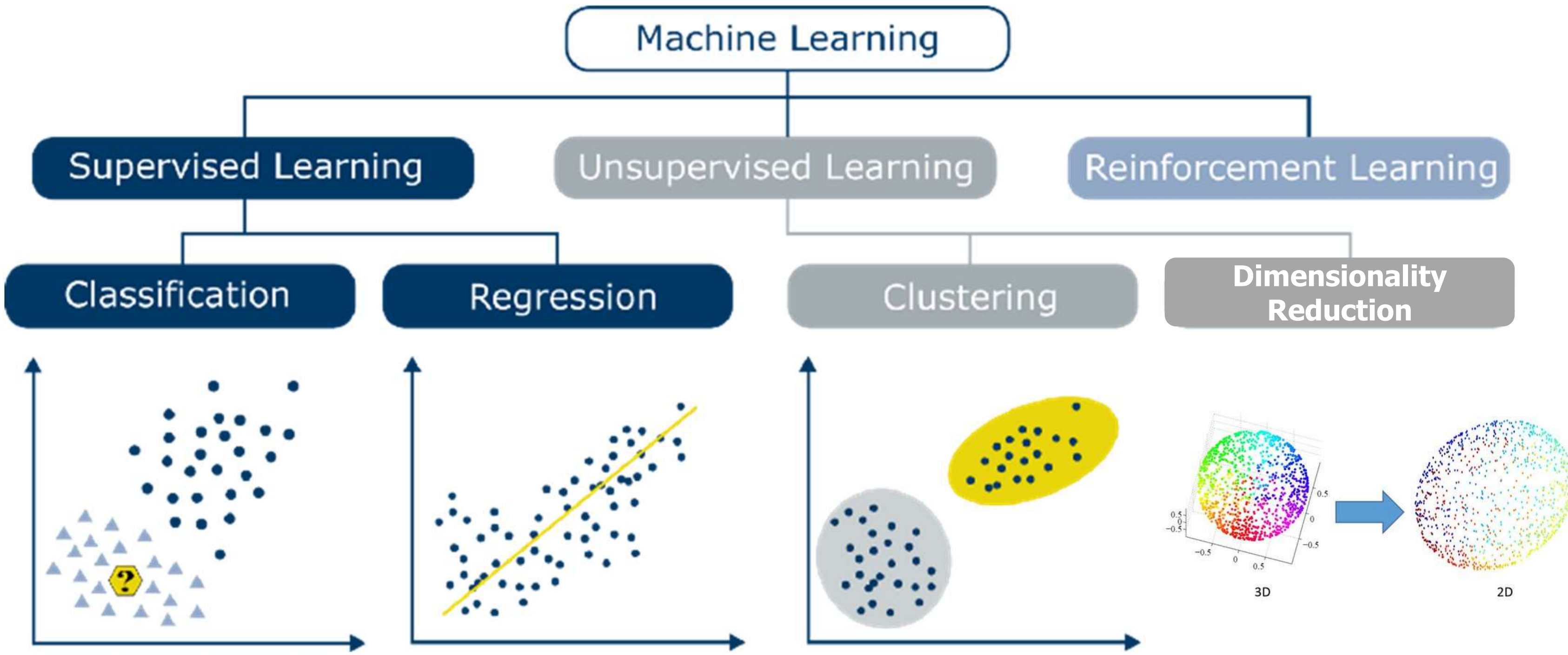
EVALUACION:
Viabilidad del negocio.
Validación criterios de éxito.

DESPLIEGUE:
Resultados.
Conocimiento.
Estrategia.

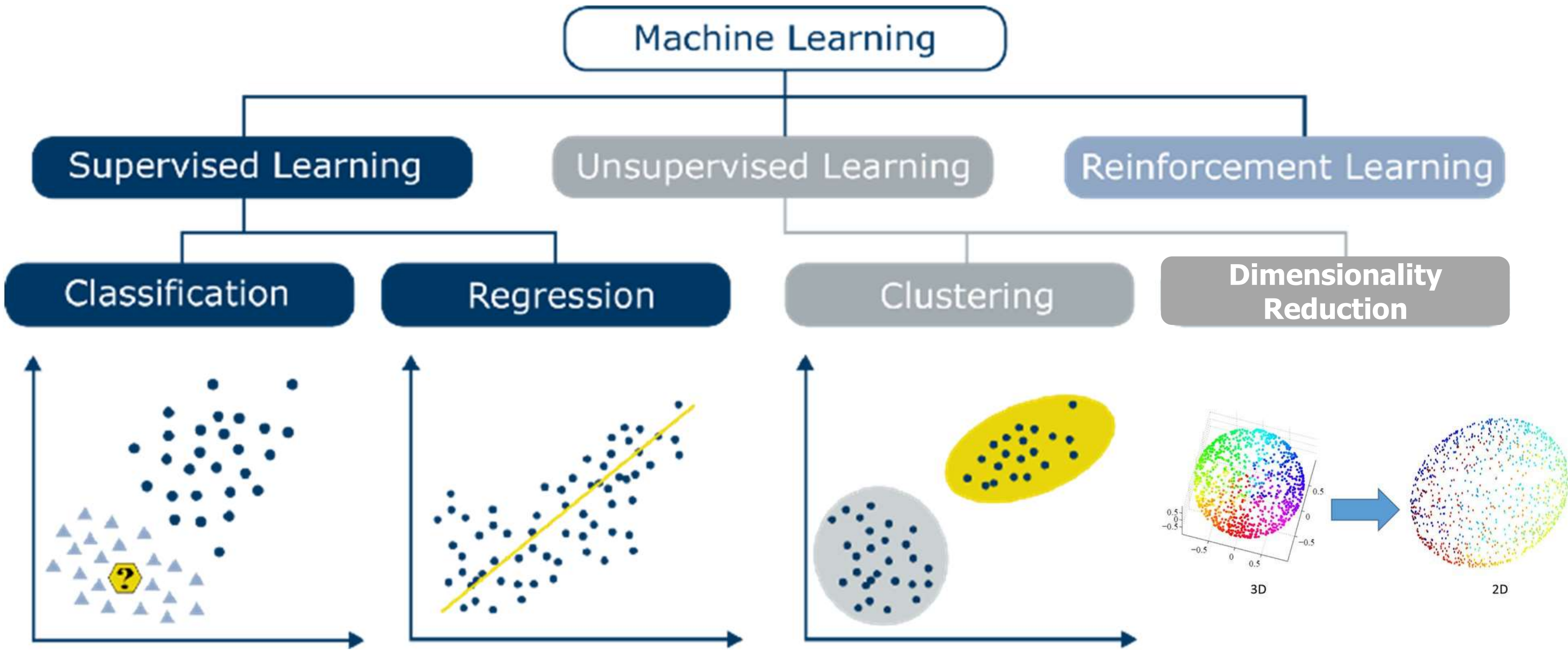
PREGUNTAS:
¿Cómo?
¿Cuáles?
¿Cuándo?
¿Por qué? *



ANALISIS EXPLORATORIO:
Limpieza de datos.
Preparación / transformación.
Selección de variables (atributos).



Métodos	• KNN • Regresión Logística	• Regresión Lineal Simple • Regresión Lineal Múltiple • Regresión Polinomial	• K-Means
Evaluación	Accuracy, Precision, Recall, F1 score, ROC Curve, AUC, etc.	MSE, RMSE, R^2, etc.	Método del codo Método de la silueta Calinski-Harabasz



Métodos

- KNN
- Regresión Logística

- Regresión Lineal Simple
- Regresión Lineal Múltiple
- Regresión Polinomial

- K-Means

- Bayes ingenuo
- Árboles de decisión
- Máquinas de soporte vectorial (SVM)
- Redes neuronales
- …

2. Repaso de probabilidad

Teoría de probabilidad

- La teoría de la probabilidad es una rama de las matemáticas que estudia los fenómenos aleatorios y estocásticos. Los fenómenos aleatorios se contraponen a los fenómenos deterministas, los cuales son resultados únicos y/o previsibles de experimentos realizados bajo las mismas condiciones determinadas, por ejemplo, si se calienta agua a 100 ºC a nivel del mar se obtendrá vapor. Los fenómenos aleatorios, por el contrario, son aquellos que se obtienen de experimentos realizados, otra vez, bajo las mismas condiciones determinadas pero como resultado posible poseen un conjunto de alternativas, por ejemplo, el lanzamiento de un dado o de una moneda.

- La teoría de la probabilidad se ocupa de asignar un cierto número a cada posible resultado que pueda ocurrir en un experimento aleatorio, con el fin de cuantificar dichos resultados y saber si un suceso es más probable que otro.

Teoría de probabilidad

- Veamos el siguiente ejemplo:

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

Teoría de probabilidad

- Veamos el siguiente ejemplo:

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

Teoría de probabilidad

- Veamos el siguiente ejemplo:

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

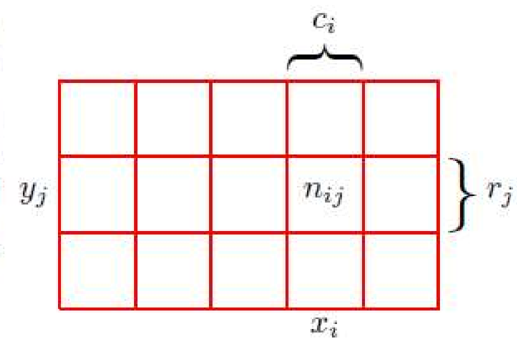
Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

Primero necesitamos comprender dos reglas fundamentales de la probabilidad: la regla de la suma y la regla del producto.

Reglas de la suma y la multiplicación

Figure 1.10 We can derive the sum and product rules of probability by considering two random variables, X, which takes the values $\{x_i\}$ where $i = 1, \ldots, M$, and Y, which takes the values $\{y_j\}$ where $j = 1, \ldots, L$. In this illustration we have $M = 5$ and $L = 3$. If we consider a total number N of instances of these variables, then we denote the number of instances where $X = x_i$ and $Y = y_j$ by n_{ij}, which is the number of points in the corresponding cell of the array. The number of points in column i, corresponding to $X = x_i$, is denoted by c_i, and the number of points in row j, corresponding to $Y = y_j$, is denoted by r_j.



Convención:

- *X*: variable aleatoria que toma valores x_i. Donde *i* varía entre 1 y *M*.
- *Y*: variable aleatoria que toma valores y_j. Donde *j* varía entre 1 y *L*.
- *N*: Número de instancias.
- n_{ij}: Número de instancias donde $X= x_i$ y $Y= y_j$.
- c_i: Número de instancias en la columna *i* donde $X= x_i$.
- r_j: Número de instancias en la fila *j* donde $Y= y_j$.

Reglas de la suma y la multiplicación

X: variable aleatoria valor Dado 1.
Y: variable aleatoria valor Dado 2.

Dado 1 / Dado 2	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(1,3)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(1,4)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(1,5)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(1,6)	(6,2)	(6,6)	(6,4)	(6,5)	(6,6)

Convención:

- X: variable aleatoria que toma valores x_i. Donde i varía entre 1 y M.
- Y: variable aleatoria que toma valores y_j. Donde j varía entre 1 y L.
- N: Número de instancias.
- n_{ij}: Número de instancias donde $X= x_i$ y $Y= y_j$.
- c_i: Número de instancias en la columna i donde $X= x_i$.
- r_j: Número de instancias en la fila j donde $Y= y_j$.

Probabilidad conjunta (*joint probability*)

- Probabilidad que X tome el valor x_i, **y** Y tome el valor y_j.

$$\boxed{p\left(X = x_i, Y = y_j\right) = \frac{n_{ij}}{N}}$$

Asumiendo el límite $N \to \infty$

Probabilidad que el Dado 1 tome el valor 6 y el Dado 2 tome el valor 3

P(X=6,Y=3) = 1/36

Dado 1 / Dado 2	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(1,3)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(1,4)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(1,5)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(1,6)	(6,2)	(6,6)	(6,4)	(6,5)	(6,6)

Probabilidad marginal (*marginal probability*)

- Probabilidad que X tome el valor x_i, independiente del valor de Y.

$$p(X = x_i) = \frac{c_i}{N} \qquad c_i = \sum_j n_{ij}$$

$$p(X = x_i) = \frac{\sum_j n_{ij}}{N} = \sum_j \frac{n_{ij}}{N} = \sum_j p(X = x_i, Y = y_j)$$

$$\boxed{p(X = x_i) = \sum_{j=1}^{L} p(X = x_i, Y = y_j)}$$

Regla de la suma

Probabilidad que el Dado 1 tome el valor 6:

P(X=6) = 6/36

También:

P(X=6) $= \sum_{j=1}^{6} p(X = 2, Y = j) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{6}{36} = \frac{1}{6}$

Dado 1 / Dado 2	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(1,3)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(1,4)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(1,5)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(1,6)	(6,2)	(6,6)	(6,4)	(6,5)	(6,6)

Probabilidad condicional (*conditional probability*)

- Si sólo se consideran aquellas instancias para las cuales $X=x_i$, entonces la fracción de estas instancias para las cuales $Y=y_j$, se denotan como:

$$p\left(Y=y_j \mid X=x_i\right)=\frac{n_{ij}}{c_i}$$

$$\boxed{p\left(Y=y_j \mid X=x_i\right)=\frac{p\left(Y=y_j, X=x_i\right)}{p\left(X=x_i\right)}}$$

- El lado izquierdo de la expresión anterior se lee como "la probabilidad que $Y=y_j$, dado que $X=x_i$".

Probabilidad que el Dado 2 tome el valor 3 dado que el Dado 1 tome el valor 6:

P(Y=3|X=6) $= \frac{p(Y=3,X=6)}{p(X=6)} = \frac{1/36}{6/36} = \frac{1}{6}$

Dado 1 / Dado 2	**1**	**2**	**3**	**4**	**5**	**6**
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(1,3)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(1,4)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(1,5)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(1,6)	(6,2)	(6,6)	(6,4)	(6,5)	(6,6)

Regla del producto

$$p\left(X = x_i, Y = y_j\right) = \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} x \frac{c_i}{N}$$

$$\boxed{p\left(X = x_i, Y = y_j\right) = p\left(Y = y_j \mid X = x_i\right) p\left(X = x_i\right)}$$

Regla del producto

Probabilidad que el Dado 1 tome el valor 6 y Dado 2 tome el valor 3:

P(X=6,Y=3) = P(Y=3|X=6)p(X=6) = $\frac{1}{6} * \frac{1}{6} = \frac{1}{36}$

Dado 1 / Dado 2	1	2	3	4	5	6
1	(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
2	(1,2)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(1,3)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(1,4)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(1,5)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(1,6)	(6,2)	(6,6)	(6,4)	(6,5)	(6,6)

Reglas de la suma y la multiplicación

Regla de la suma:	$p(X) = \sum_{Y} p(X,Y)$
Regla del producto:	$p(X,Y) = p(Y \mid X)\, p(X)$

Estas dos reglas forman las bases para toda la maquinaria probabilística que se usa en reconocimiento de patrones.

Ejemplo de una distribución sobre dos variables

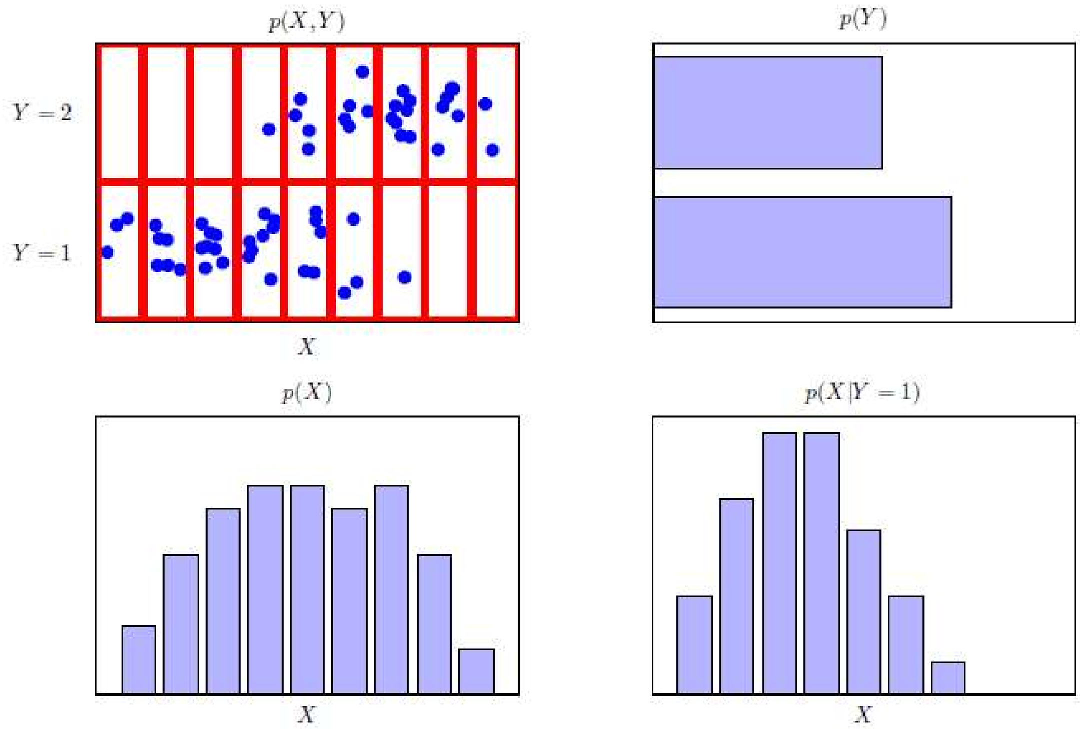


Figure 1.11 An illustration of a distribution over two variables, X, which takes 9 possible values, and Y, which takes two possible values. The top left figure shows a sample of 60 points drawn from a joint probability distribution over these variables. The remaining figures show histogram estimates of the marginal distributions $p(X)$ and $p(Y)$, as well as the conditional distribution $p(X|Y=1)$ corresponding to the bottom row in the top left figure.

Teorema de Bayes

- Usando la *regla del producto* y la *regla de simetría* $p(X,Y)=p(Y,X)$, se obtiene la siguiente relación entre probabilidades condicionales:

$$\boxed{p(Y \mid X) = \frac{p(X \mid Y)\,p(Y)}{p(X)}} \qquad \boxed{p(X \mid Y) = \frac{p(Y \mid X)\,p(X)}{p(Y)}}$$

Usando $p(X) = \sum_Y p(X \mid Y)\,p(Y)$ $\qquad p(Y) = \sum_X p(Y \mid X)\,p(X)$

El teorema de Bayes se puede reescribir como:

$$\boxed{p(Y \mid X) = \frac{p(X \mid Y)\,p(Y)}{\sum_Y p(X \mid Y)\,p(Y)}} \qquad \boxed{p(X \mid Y) = \frac{p(Y \mid X)\,p(X)}{\sum_X p(Y \mid X)\,p(X)}}$$

Retomando el ejemplo

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

X = estar sano o no sano.
Y = prueba positiva o prueba negativa.

$$P(X=1|Y=1)=\frac{P(Y=1|X=1)P(X=1)}{P(Y=1)}$$

Retomando el ejemplo

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

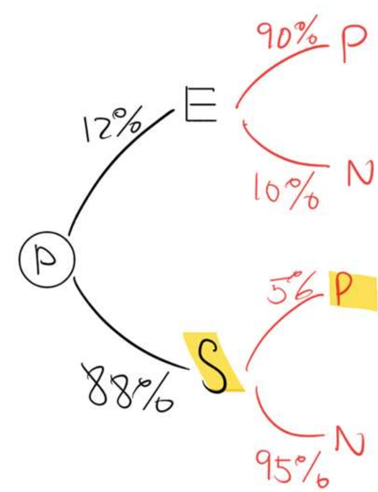
Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

X = estar sano o no sano.
Y = prueba positiva o prueba negativa.

$$P(X=1|Y=1) = \frac{P(Y=1|X=1)P(X=1)}{P(Y=1)}$$

$$P(X=1|Y=1) = \frac{0.05 * 0.88}{P(Y=1)}$$



Retomando el ejemplo

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

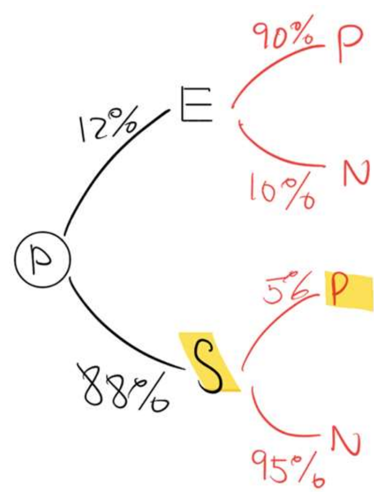
Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

X = estar sano o no sano.
Y = prueba positiva o prueba negativa.

$$P(X = 1|Y = 1) = \frac{P(Y = 1|X = 1)P(X = 1)}{P(Y = 1)}$$

$$P(X = 1|Y = 1) = \frac{0.05 * 0.88}{P(Y = 1)}$$



$$p(Y = 1) = \sum_j p(X = j, Y = 1)$$

Retomando el ejemplo

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

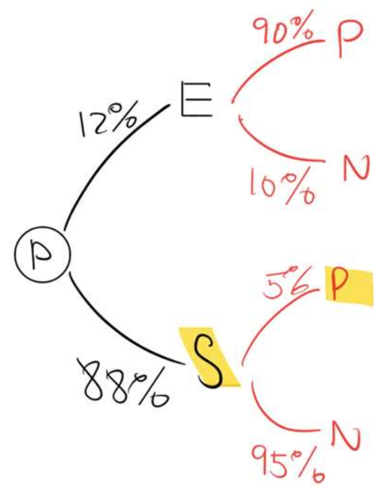
Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

X = estar sano o no sano.
Y = prueba positiva o prueba negativa.

$$P(X=1|Y=1) = \frac{P(Y=1|X=1)P(X=1)}{P(Y=1)}$$

$$P(X=1|Y=1) = \frac{0.05 * 0.88}{P(Y=1)}$$



$$p(Y=1) = \sum_j p(X=j, Y=1)$$

$$= P(X=0, Y=1) + P(X=1, Y=1)$$

$$= (0.12) * (0.9) + (0.88) + (0.05)$$

$$= 0.152$$

Retomando el ejemplo

En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

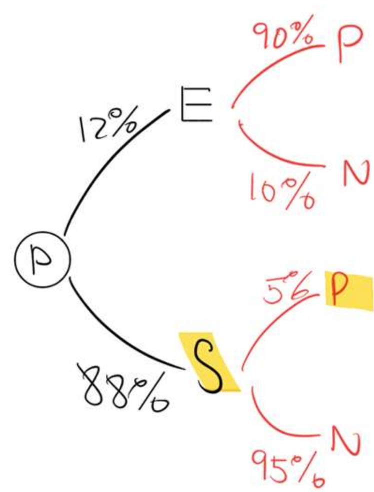
Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

X = estar sano o no sano.
Y = prueba positiva o prueba negativa.

$$P(X=1|Y=1) = \frac{P(Y=1|X=1)P(X=1)}{P(Y=1)}$$

$$P(X=1|Y=1) = \frac{0.05 * 0.88}{0.152} = 0.2894$$



Retomando el ejemplo

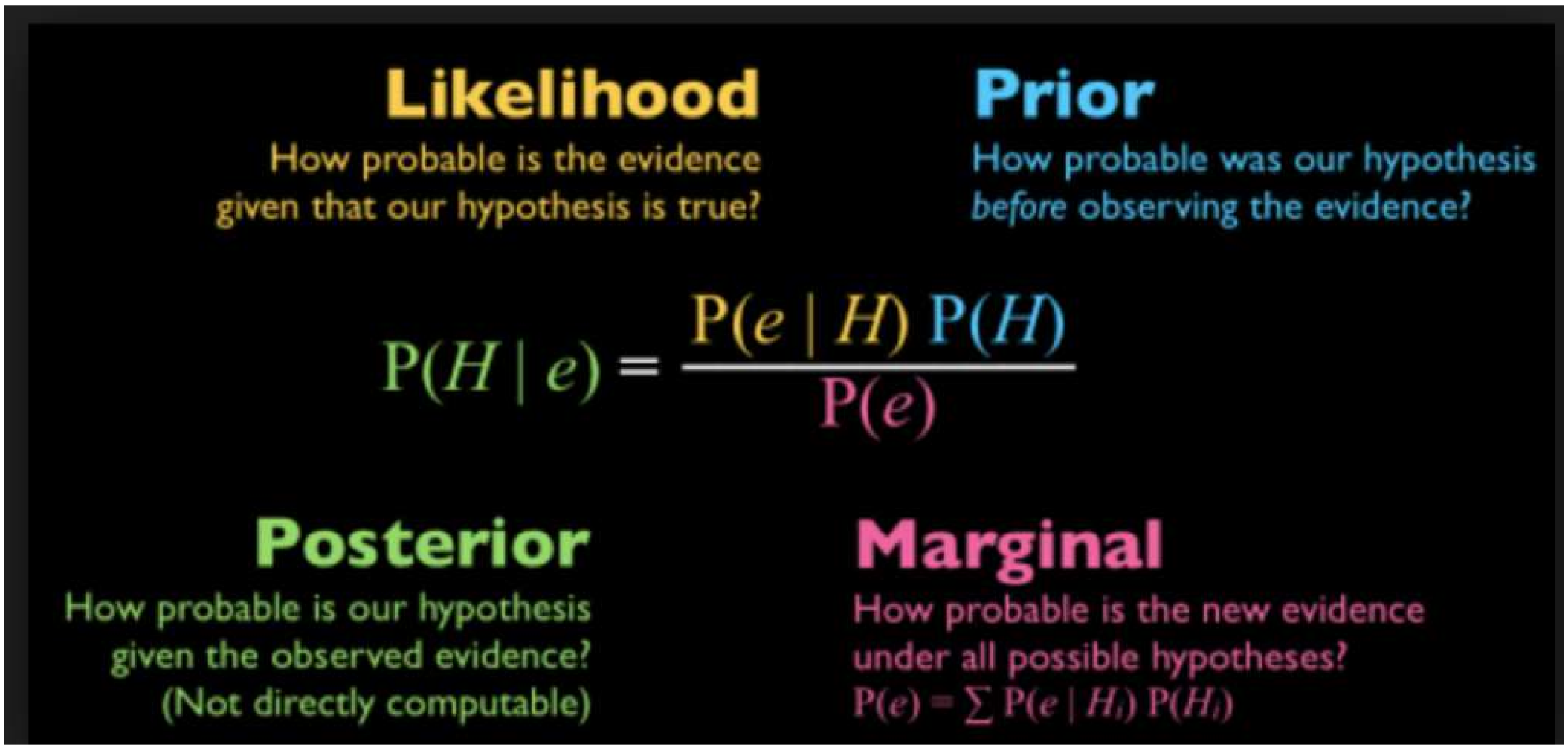
En cierto país donde la enfermedad X es endémica, se sabe que un12% de la población padece dicha enfermedad.

Se dispone de una prueba para detectar la enfermedad, pero no es totalmente fiable, ya que, da positivo en el 90% de personas realmente enfermas y da positivo en el 5% de personas sanas.

¿Cuál es la probabilidad de que esté sana una persona a la que la prueba le ha dado positiva?

$\mathbf{28.94}\%$

¿Por qué es útil el teorema de Bayes?



Independencia

- Si la probabilidad conjunta:

$$p(X, Y) = p(X)p(Y)$$

- Entonces se puede decir que X y Y son eventos independientes y por lo tanto la probabilidad condicional se puede expresar como:

$$p(Y|X) = p(Y)$$

3. Clasificador Bayesiano Ingenuo (Naïve Bayes classifier)

Nota: Los métodos Bayesianos ingenuos son un conjunto de algoritmos de aprendizaje supervisados basados en la aplicación del teorema de Bayes con el supuesto "*ingenuo*" de *independencia condicional* entre cada par de características dado el valor de la variable de clase $(ejemplo: x_1 \perp x_2 \mid Y)$.

Clasificadores Bayesianos

- Los clasificadores Bayesianos asignan cada observación a la clase j más probable, dados los valores observados de sus variables predictivas, es decir:

$$argmax_{j}\ p(Y=y_j|X=x_{observados}).$$

- Si se conocen las distribuciones de probabilidad condicional, el clasificador resultante da la frontera de separación óptima en términos de error. Sin embargo, este no siempre es el caso.

- Algunas aplicaciones de los clasificadores Bayesianos son: clasificación de texto (e.g.: *spam* vs. *no spam*), diagnóstico de enfermedades a partir de un conjunto de síntomas, etc.

Clasificador Bayesiano Ingenuo

Regla de Bayes:

$$\overbrace{p\left(y_j \middle| x_1, x_2, \dots, x_n\right)}^{\text{Probabilidad Posterior}} = \frac{p(y_j, x_1, x_2, \dots, x_n)}{p(x_1, x_2, \dots, x_n)} = \frac{\overbrace{p(y_j)}^{\text{Probabilidad A priori}} * \overbrace{p(x_1, x_2, \dots, x_n | y_j)}^{\text{Verosimilitud}}}{\underbrace{p(x_1, x_2, \dots, x_n)}_{\text{Evidencia}}}$$

Note que el denominador sólo se usa para propósitos de normalización (i.e.: que la suma de probabilidades sea 1), pues:

$$p(x_1, x_2, \dots, x_n) = \sum_j \left(p(y_j) \; * \; p(x_1, x_2, \dots, x_n | y_j) \right)$$

Por ello, sólo nos interesa el numerador:

$$p\left(y_j, x_1, x_2, \dots, x_n\right) = p\left(y_j\right) * p\left(x_1 \middle| y_j\right) * p\left(x_2 \middle| x_1, y_j\right) * p\left(x_3 \middle| x_2, x_1, y_j\right) * \cdots * p\left(x_n \middle| x_{1:n-1}, y_j\right)$$

Luego, si asumimos ingenuamente (*naïvely*) que todas las variables predictivas x_i son condicionalmente independientes dada la clase y_j, entonces el numerador se puede simplificar como:

$$p\left(y_j\right) * p\left(x_1 \middle| y_j\right) * p\left(x_2 \middle| y_j\right) * p\left(x_3 \middle| y_j\right) * \cdots * p\left(x_n \middle| y_j\right)$$

$$= p(y_j) \prod_{i=1}^{n} p\left(x_i \middle| y_j\right)$$

Clasificador Bayesiano Ingenuo

La regla de clasificación es:

$$arg \max_j [p(y_j) * \prod_{i=1}^{n} p(x_i|y_j)]$$

Sólo necesitamos especificar:

- Las probabilidades a priori de cada clase.
- Las **distribuciones** de probabilidad condicional de las variables predictivas para cada clase (i.e.: distribuciones de probabilidad condicionadas a la clase).

Esta información constituye los **parámetros** del modelo, y en el caso de variables categóricas se obtienen a partir de frecuencias (conteos).

Clasificador Bayesiano Ingenuo

Teniendo en cuenta el tipo de datos de los atributos de entrada es necesario utilizar la **distribución** adecuada $p(x_i|y_j)$:

- **Bernoulli**: se utiliza para las variables **binarias.**
- **Multinomial**: se utiliza para las variables con **conteos.**
- **Gaussiana**: se utiliza para variables cuantitativas **continuas**, en las que se asume una distribución **normal**.
- **Kernels**: se utiliza para variables cuantitativas **continuas**, en las que **no se asume una distribución paramétrica**. En este caso el "kernel" es una función de probabilidad que se establece dinámicamente dados los valores de la variable en cuestión.

Clasificador Bayesiano Ingenuo

- Es posible que con algunos de los valores de las variables predictivas tengan frecuencia nula con respecto a las categorías de la clase en los datos de entrenamiento, por lo cual, sus probabilidades asociadas serían cero.
- Para evitar este problema, se utilizan métodos de **suavización**, que agregan un valor pequeño $\boldsymbol{\varepsilon}$ al contar las frecuencias de ocurrencia de cada valor, esto impide las probabilidades puedan sean cero:

$$P(casado|cliente\ potencial) = \frac{Conteo(casado,\ cliente\ potencial) + \varepsilon}{Conteo(cliente\ potencial) + n * \varepsilon}$$

Donde: "n" representa el número de atributos en los datos de entrada.

- El método de suavización de **Laplace** se aplica usualmente con ε=1, utilizando distribuciones Bernouilli o multinomiales.

Clasificador Bayesiano Ingenuo

Cuando las variables predictivas no son categóricas, es necesario establecer una distribución de probabilidad:

1. Se puede discretizar la variable convirtiéndola en categórica.
2. Se puede establecer una distribución de probabilidad empírica utilizando KNN.
3. Se puede suponer eventualmente que se trata de un tipo de distribución de probabilidad y utilizar su función de densidad.

Por ejemplo, si se supone que se trata de una variable que sigue una distribución normal condicionada a la categoría objetivo, se puede calcular la media μ y desviación estándar σ a partir de los datos históricos, y utilizar la función de densidad. Por ejemplo:

$$P(edad|cliente\ potencial) = \frac{1}{\sigma_{edad|cliente}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{edad-\mu_{edad|cliente}}{\sigma_{edad|cliente}}\right)^2}$$

Clasificador Bayesiano Ingenuo

Consideraciones

- Sólo se puede utilizar para **clasificación.**
- Es un modelo **simple** y **eficiente para** atributos categóricos (dos o más) o numéricos.
- El clasificador tolera atributos con **valores faltantes.**
- Ignora atributos **irrelevantes.**
- **Es muy sensible** a atributos correlacionados (pues considera varias veces los mismos efectos).
- Es resistente al sobre ajuste, especialmente si se incluye el suavizado (e.g.: Laplace).
- Es ideal cuando se tienen dimensionalidades altas en los datos de entrada.

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(y_j) \prod_{i=1}^{n} p\left(x_i \middle| y_j\right)$$

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(\boxed{y_j}) \prod_{i=1}^{n} p(x_i | y_j)$$

Subscribed=yes	11%		Subscribed=no	88%

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(\boxed{y_j}) \prod_{i=1}^{n} p(\boxed{x_i}|y_j)$$

Subscribed=yes	11%

Subscribed=no	88%

Atributos:
x1=Job
x2=Marital
x3=Education
x4=Default
x5=Housing
x6=Loan
x7=Contact
x8=Outcome

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(\boxed{y_j}) \prod_{i=1}^{n} p(\boxed{x_i}|y_j)$$

Subscribed=yes	11%

Subscribed=no	88%

Atributos:
x1=Job
x2=Marital
x3=Education
x4=Default
x5=Housing
x6=Loan
x7=Contact
x8=Outcome

{Single, Married, Divorced}

{Subscribed Yes, Subscribed No}

$$p(\boxed{x_i}|\boxed{y_j})$$

Marital	Subscribed=yes
Single	35%
Married	53%
Divorced	12%

Marital	Subscribed=no
Single	28%
Married	61%
Divorced	11%

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(y_j) \prod_{i=1}^{n} p(x_i | y_j)$$

¿Debería el banco ofrecerle un CDT al cliente con la siguiente información?

	Atributos
x1	Job=Management
x2	Marital=Married
x3	Education=Secondary
x4	Default=no
x5	Housing=yes
x6	Loan=no
x7	Contact=Cellular
x8	Outcome=Success

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(y_j) \prod_{i=1}^{n} p(x_i | y_j)$$

¿Debería el banco ofrecerle un CDT al cliente con la siguiente información?

	Atributos
x1	Job=Management
x2	Marital=Married
x3	Education=Secondary
x4	Default=no
x5	Housing=yes
x6	Loan=no
x7	Contact=Cellular
x8	Outcome=Success

Suponga que se disponen de las probabilidades condicionales para todas las variables predictivas (ya ilustradas para el estado civil ("*Marital*" en Inglés))

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(y_j) \prod_{i=1}^{n} p(x_i | y_j)$$

¿Debería el banco ofrecerle un CDT al cliente con la siguiente información?

	Atributos	Subscribed=yes	Subscribed=no
x1	Job=Management	22%	21%
x2	Marital=Married	53%	61%
x3	Education=Secondary	46%	51%
x4	Default=no	99%	98%
x5	Housing=yes	35%	57%
x6	Loan=no	90%	85%
x7	Contact=Cellular	85%	62%
x8	Outcome=Success	15%	1%
	Priors	11%	88%

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(y_j) \prod_{i=1}^{n} p(x_i|y_j)$$

Numerador(yes) = 0,11*0,22*0,53*0,46*0,99*0,35*0,90*0,85*0,15
Numerador(yes) = 0,000234588

Numerador(no) = 0,11*0,21*0,61*0,51*0,98*0,57*0,85*0,62*0,01
Numerador(no) = 0,000169244

p(suscribed=yes) = Numerador(yes) /(Numerador(yes) + Numerador(no)) = 58%
p(suscribed=no) = Numerador(no) /(Numerador(yes) + Numerador(no)) = 42%

¿Debería el banco ofrecerle un CDT al cliente con la siguiente información?

	Atributos	Subscribed=yes	Subscribed=no
x1	Job=Management	22%	21%
x2	Marital=Married	53%	61%
x3	Education=Secondary	46%	51%
x4	Default=no	99%	98%
x5	Housing=yes	35%	57%
x6	Loan=no	90%	85%
x7	Contact=Cellular	85%	62%
x8	Outcome=Success	15%	1%
	Priors	11%	88%
	Numerador	0,000234588	0,000169244
	Probabilidad posterior	58%	42%

Clasificador Bayesiano Ingenuo

Ejemplo: Un banco quiere predecir si un cliente va a adquirir un CDT.

Creamos un clasificador Bayesiano ingenuo a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: *subscribed=yes* y *subscribed=no*.

$$argmax_j \; p(y_j) \prod_{i=1}^{n} p(x_i | y_j)$$

Numerador(yes) = 0,11*0,22*0,53*0,46*0,99*0,35*0,90*0,85*0,15
Numerador(yes) = 0,000234588

Numerador(no) = 0,11*0,21*0,61*0,51*0,98*0,57*0,85*0,62*0,01
Numerador(no) = 0,000169244

p(suscribed=yes) = Numerador(yes) /(Numerador(yes) + Numerador(no)) = 58%
p(suscribed=no) = Numerador(no) /(Numerador(yes) + Numerador(no)) = 42%

¿Debería el banco ofrecerle un CDT al cliente con la siguiente información?

	Atributos	Subscribed=yes	Subscribed=no
x1	Job=Management	22%	21%
x2	Marital=Married	53%	61%
x3	Education=Secondary	46%	51%
x4	Default=no	99%	98%
x5	Housing=yes	35%	57%
x6	Loan=no	90%	85%
x7	Contact=Cellular	85%	62%
x8	Outcome=Success	15%	1%
	Priors	11%	88%

Numerador	0,000234588	0,000169244
Probabilidad posterior	58%	42%

Pregunta: ¿Se le debe ofrecer el CDT al cliente?

¿Cómo invocar el clasificador Bayesiano Ingenuo desde Scikit-learn en Python?

- *Gaussian Naive Bayes*:
  - ✓ from sklearn.naive_bayes import GaussianNB

- *Multinomial Naive Bayes*:
  - ✓ from sklearn.naive_bayes import MultinomialNB

- *Bernoulli Naive Bayes*:
  - ✓ from sklearn.naive_bayes import BernoulliNB

- *Categorical Naive Bayes*:
  - ✓ from sklearn.naive_bayes import CategoricalNB

- Para mayor información, puede consultar: https://scikit-learn.org/stable/modules/naive_bayes.html

¿Se puede usara el clasificador Bayesiano ingenuo si los tipos de datos de entrada son mixtos?

- Si, pero con la implementación actual de Scikit-learn debe hacerse por partes. Para ver los detalles consulte el siguiente blog: *Naive-Bayes for mixed typed data in scikit-learn*

Lecturas Complementarias

- Naive Bayes Classifier
(https://www.javatpoint.com/machine-learning-naive-bayes-classifier)

- Algoritmo Naïve Bayes: Fundamentos e Implementación
(https://medium.com/datos-y-ciencia/algoritmos-naive-bayes-fudamentos-e-implementaci%C3%B3n-4bcb24b307f)