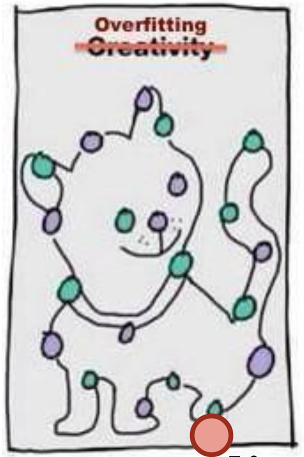
APRENDIZAJE SUPERVISADO

Javier Diaz Cely, PhD

AGENDA

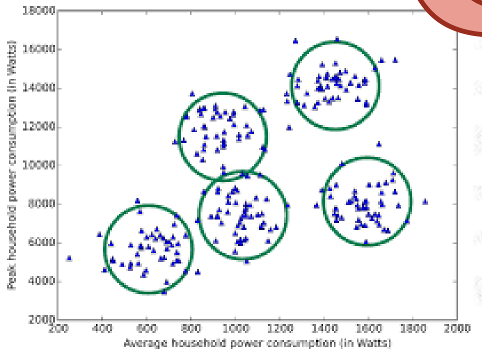
Protocolos



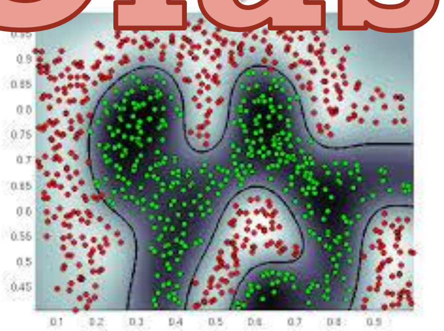
Aprendizaje automático

Sobreaprendizaje (Overfitting)

Clase anterior

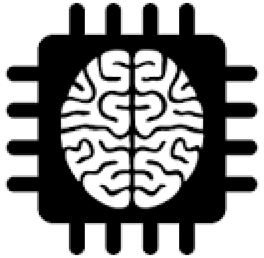


Aprendizaje no supervisado

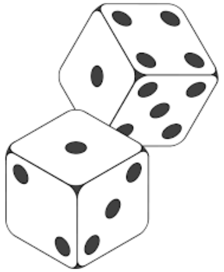


Aprendizaje supervisado

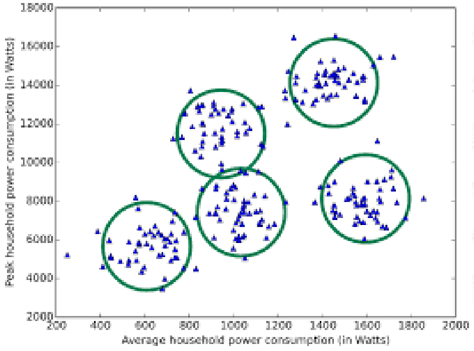
AGENDA



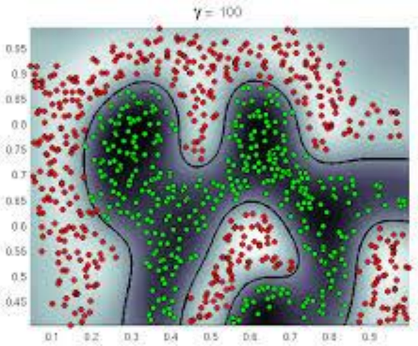
Aprendizaje automático



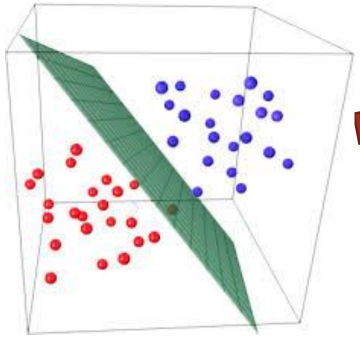
Probabilidad



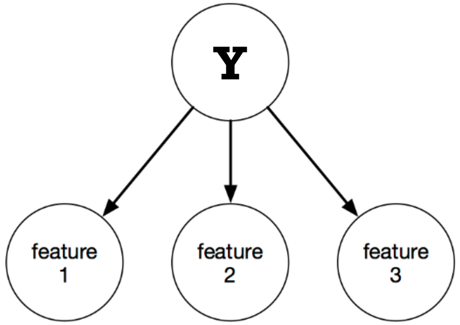
Aprendizaje no supervisado



Aprendizaje supervisado

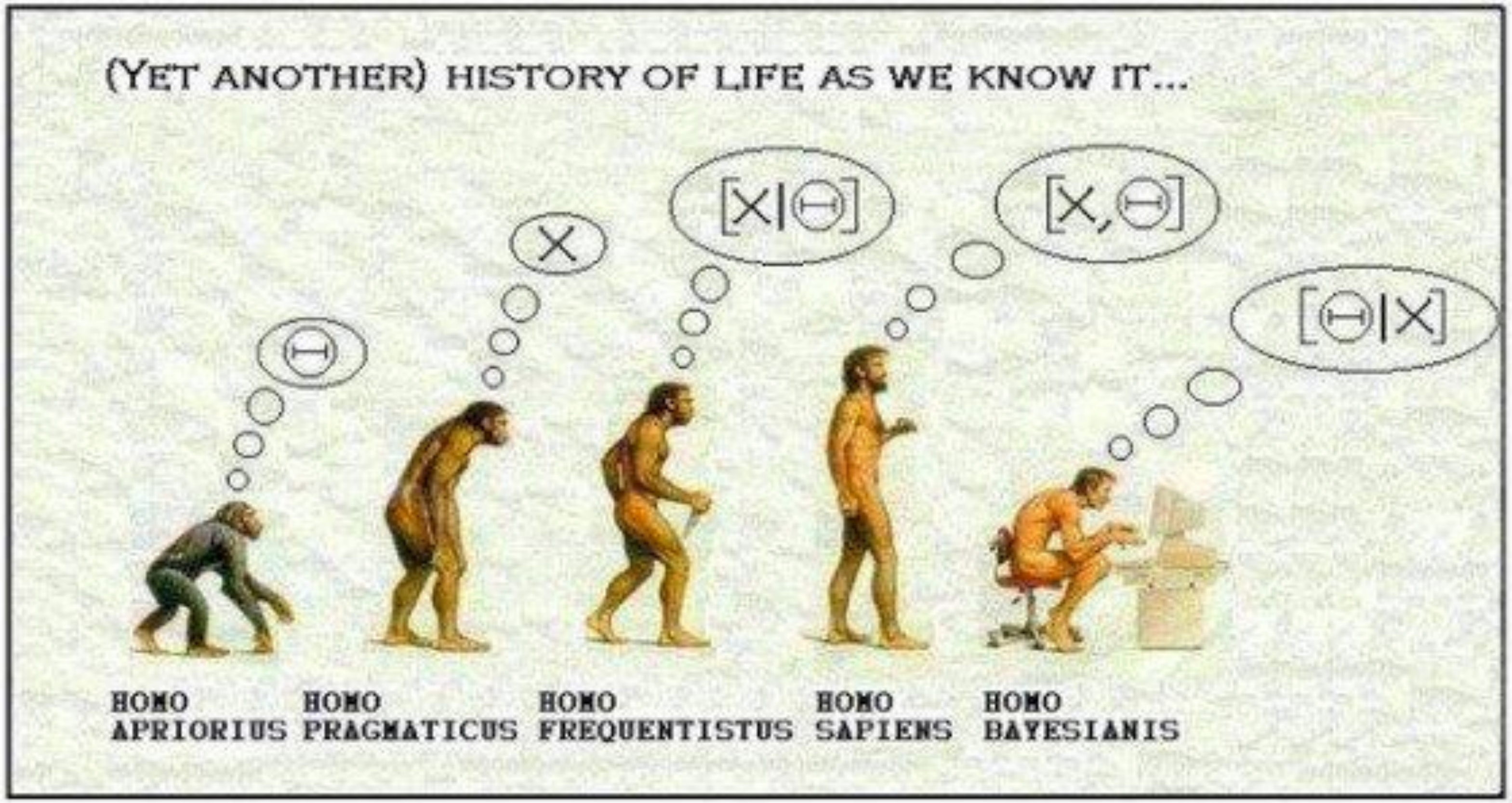


Clasificación



Naïve Bayes

NAÏVE BAYES


(YET ANOTHER) HISTORY OF LIFE AS WE KNOW IT...
[Θ]
[X]
[X|Θ]
[X,Θ]
[Θ|X]
HOMO APRIORIUS
HOMO PRAGMATICUS
HOMO FREQUENTISTUS
HOMO SAPIENS
HOMO BAYESIANIS

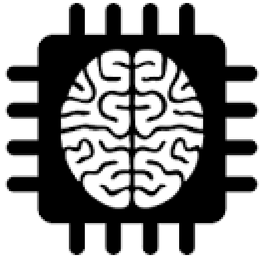
NAIVE BAYES: TALLER

1. Socrative quiz probabilidades

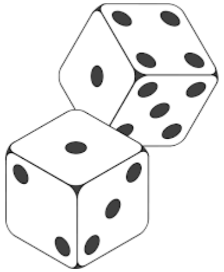
2. Descarguen el taller de Excel y Word de Naive Bayes.

Desarrollen las partes 1 y 2 del taller, de repaso del calculo de probabilidades básicas y de entendimiento de la condicionalidad

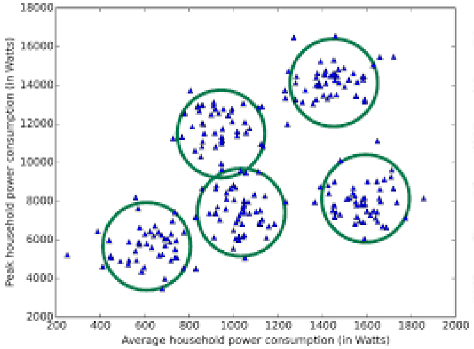
AGENDA



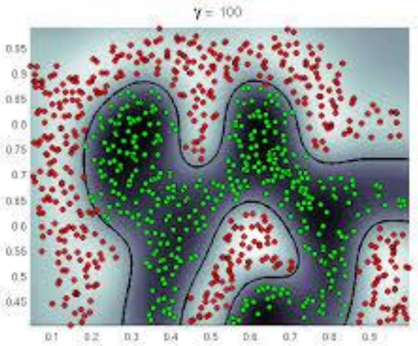
Aprendizaje automático



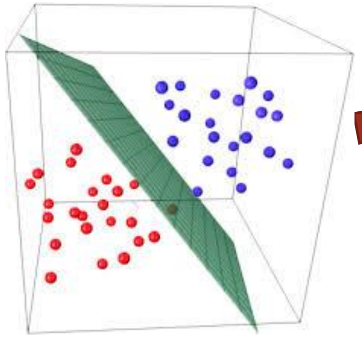
Probabilidad



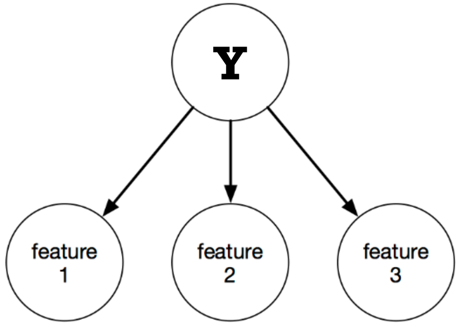
Aprendizaje no supervisado



Aprendizaje supervisado



Clasificación



Naïve Bayes

PROBABILIDADES

Marginalización: $p(X = x_i) = \sum_j p(x_i, y_j)$

Regla de producto: $p(X = x_i, Y = y_j) = p(Y = y_j | X = x_i) * p(X = x_i)$

$p(X = x_i, Y = y_j) = p(X = x_i | Y = y_j) * p(Y = y_j)$

Regla de Bayes: $p(Y = y_j | X = x_i) = \frac{p(X = x_i | Y = y_j) * p(Y = y_j)}{p(X = x_i)}$

Independencia: $p(Y = y_j | X = x_i) = p(Y = y_j)$

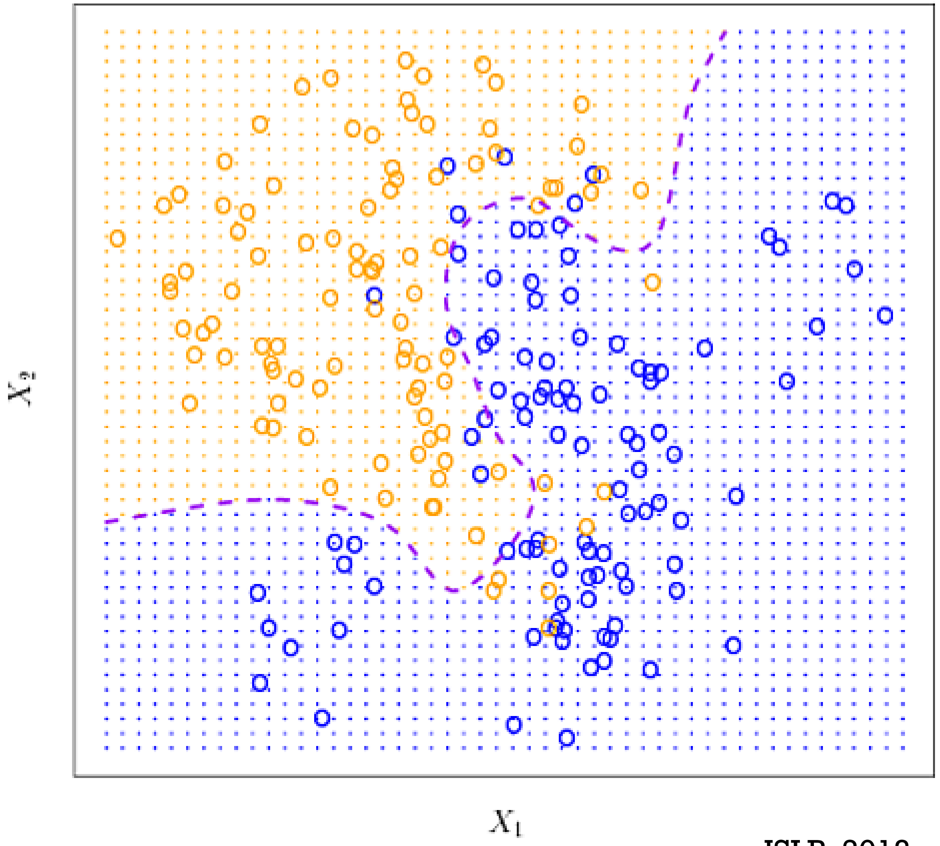
$p(X = x_i, Y = y_j) = p(X = x_i) * p(Y = y_j)$

CLASIFICADORES BAYESIANOS

- **Clasificadores bayesianos**: Asignar cada observación a la clase *j* más probable, dados los valores observados de sus variables predictivas:

$$argmax_j \; p(Y = y_j | X = x_{observados})$$

- Si se conoce perfectamente las distribuciones de probabilidad, el clasificador resultante da la frontera de separación óptima en términos de error
- No siempre se tienen las probabilidades condicionales necesarias.
- **Naïve Bayes** propone una simplificación



ISLR, 2013

NAIVE BAYES

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

Creamos un clasificador Naïve Bayes a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: subscribed=yes and subscribed=no.

{Single, Married, Divorced}

{Subscribed Yes, Subscribed No}

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

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

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

Subscribed=yes	11%

Subscribed=no	88%

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

Job=Management
Marital=Married
Education=Secondary
Default=no
Housing=yes
Loan=no
Contact=Cellular
Outcome=Success

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

NAIVE BAYES

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

Creamos un clasificador Naïve Bayes a partir de los datos históricos para calcular las probabilidades posteriores para cada clase: subscribed=yes and subscribed=no.

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

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

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

Subscribed=yes	11%

Subscribed=no	88%

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

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

Numerador	0.000234588	0.000169244
Proba posterior	58%	42%

NAIVE BAYES (BAYES INGENUO)

Probabilidad Posterior

Probabilidad A priori

Verosimilitud

Regla de Bayes: $$p(y_j|x_1, x_2, \dots, x_n) = \frac{p(y_j, x_1, x_2, \dots, x_n)}{p(x_1, x_2, \dots, x_n)} = \frac{p(y_j) * p(x_1, x_2, \dots, x_n|y_j)}{p(x_1, x_2, \dots, x_n)}$$

evidencia

El denominador es solo usado para propósitos de normalización (suma de probabilidades = 1)

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

Solo nos interesa el numerador:

$$p(y_j, x_1, x_2, \dots, x_n) = p(y_j) * p(x_1|y_j) * p(x_2|x_1, y_j) * p(x_3|x_2, x_1, y_j) * \cdots * p(x_D|x_{1:D-1}, y_j)$$

Si asumimos ingenuamente (**naïvely**) que todas las variables predictivas x_i son independientes condicionalmente con respecto a la clase y_j, entonces el numerador se simplifica:

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

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

NAÏVE BAYES (BAYES INGENUO)

La regla de clasificación es:

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

Sólo necesitamos especificar

→ Las probabilidades a priori de cada clase

→ Las distribuciones de probabilidad de las variables predictivas para cada clase (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)

NAIVE BAYES: TALLER EXCEL

Taller de Excel de Naive Bayes.

Continuar con la parte 3, aplicando Bayes ingenuo para dos variables predictivas categóricas

NAÏVE BAYES (BAYES INGENUO)

Es posible que con algunos de los valores de las variables predictivas tengan frecuencia nula con respecto a las categorías de la clase, por lo sus probabilidades asociadas serían cero.

Para evitar este problema, se utilizan métodos de **suavización**, que al contar las frecuencias de ocurrencia de cada valor, siempre se le agrega un valor pequeño ε, que impide que alguna probabilidad sea cero:

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

El método de suavización de **Laplace** se aplica con $\varepsilon=1$

NAÏVE BAYES (BAYES 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:

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

NAÏVE BAYES (BAYES INGENUO)

Consideraciones

- Sólo se puede utilizar para **clasificación**
- Modelo **simple** y **eficiente**, que permite atributos tanto categóricos (2 o más) como numéricos,
- Sólo se necesita poder estimar **las probabilidades condicionales** con respecto a los valores de la categoría objetivo, pero se basa en **suposiciones** muy fuertes (aunque en la práctica obtiene resultados buenos en muchos contextos)
- Permite atributos con **valores faltantes**
- Ignora atributos **irrelevantes**
- **Muy sensible** a atributos correlacionados (considerar varias veces los mismos efectos)
- Resistente al **overfitting**, sobretodo si se incluye un suavizador (e.g. Laplace)
- Ideal cuando se tiene un gran número de dimensiones

NAIVE BAYES: TALLER EXCEL

Taller de Excel de Naive Bayes.

Continuar con la parte 4, aplicando Bayes ingenuo con una combinación de variables categóricas y numéricas.

NAIVE BAYES: IRIS

Taller de Python de Naive Bayes aplicado al dataset Iris.

REFERENCIAS

- *Introduction to Statistical Learning with Applications in R (ISLR),* G. James, D. Witten, T. Hastie & R. Tibshirani, 2014
- *Machine Learning with R*, Brett Lantz, Packt Publishing, 2015
- *Machine Learning,* Tom M. Mitchell, McGraw-Hill, 1997
- *Real World Machine Learning*, Henrik Brink, Joseph W. Richards, Mark Fetherolf, 2017