SECME 2018

Sistema para Evaluar
la Calidad de Medios Educativos

Universidad Autónoma del Estado de México
Centro Universitario UAEM Valle de Chalco
MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN

Fundamentos de la Minería de Datos

UNIDAD DE APRENDIZAJE:
MINERÍA DE DATOS

Presenta:
DR. JOSÉ LUIS SÁNCHEZ RAMÍREZ

CONVOCATORIA 2018

PROGRAMA DE LA U.A.

I. Introducción

II. Métodos para el tratamiento y análisis de datos

III. Proceso de análisis supervisado

IV. Proceso de análisis no supervisado

V. Métodos estimadores de error

VI. Métodos para análisis del índice de acierto.

OBJETIVO DE LA UNIDAD DE APRENDIZAJE

- Al término del curso el estudiante conocerá y aplicara las metodologías para la predicción de datos que permitan pronosticar salidas de datos y revelar sus relaciones a partir de algoritmos empleados en la minería de datos: supervisados y no supervisados.

Descripción del Material

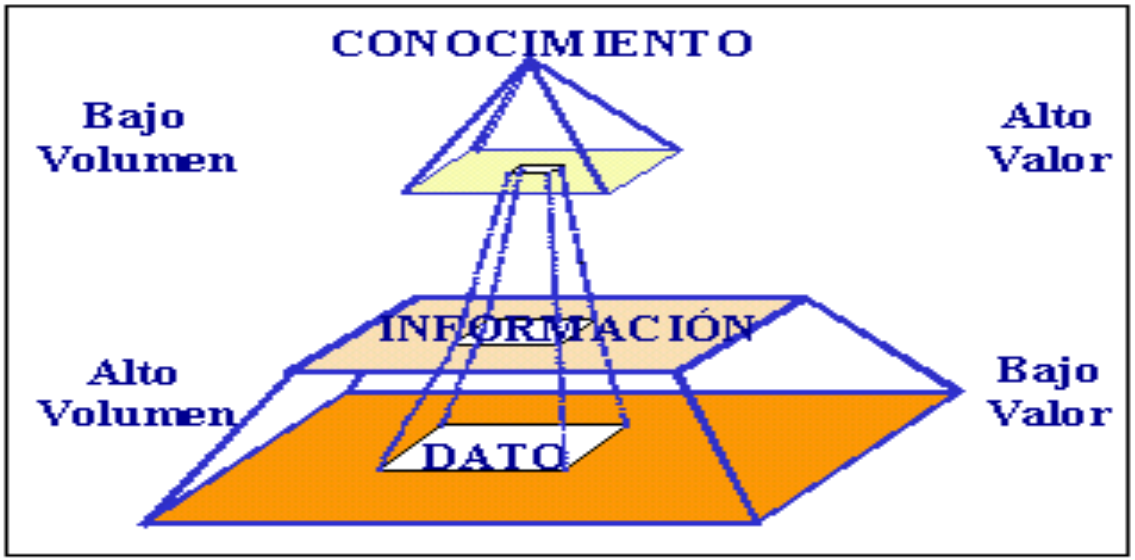
- Esta presentación esta desarrollada con base a la unidad de aprendizaje (UA) de **Minería de Datos** (MD) del programa de estudios de la *Maestría en Ciencias de la Computación* (MACSCO) **como un apoyo para abordar los contenidos** del curso teórico-práctico e ir desarrollando los temas durante el transcurso del semestre.

- Se aborda el 100% del contenido del temario de la UA a manera introductoria, sin embargo el curso se complementará y profundizará con la práctica mediante el desarrollo e implementación de los algoritmos de MD en el Software Licenciado Matlab, así como en el Software Libre WEKA.

UNIDAD I:
INTRODUCCIÓN

Introducción

Día a día generamos información y esto nos lleva a tener una gran cantidad de esta, lo cual implica que el generar información, nos puede ayudar a controlar, optimizar, administrar, examinar, investigar, planificar, predecir, someter, negociar o tomar decisiones de cualquier ámbito según el dominio en que nos desarrollemos.

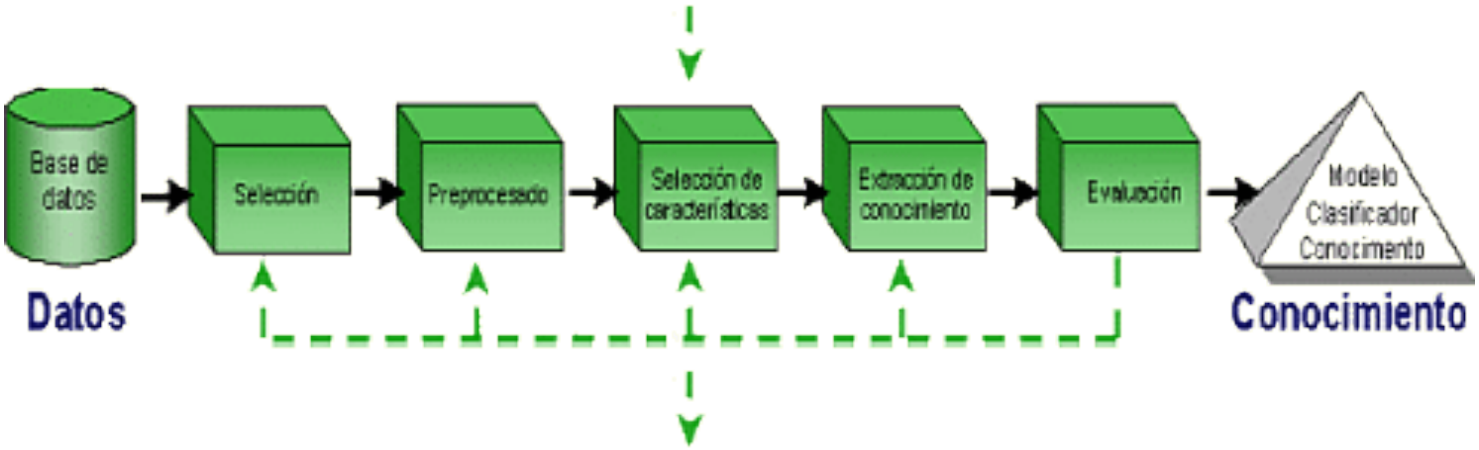


¿Qué es Minería de Datos?

- La extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de datos. (1)

- La integración de un conjunto de áreas que tienen como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten un sesgo hacia la toma de decisión(2)

Proceso de Minería de Datos 1/2



Los pasos a seguir para la realización de un proyecto de minería de datos son:

1. La Determinación de los Objetivos. Trata sobre la delimitación de los objetivos que se requieran

2. Pre procesamiento de los Datos. Se refiere a la selección, la limpieza, el enriquecimiento, la reducción y transformación de las bases de datos.

Proceso de Minería de Datos 2/2

3. Determinación del Modelo. Se comienza realizando unos análisis estadísticos de los datos y después se lleva a cabo una visualización gráfica de los mismos para tener una primera aproximación. Según los objetivos planteados y la tarea que debe llevarse a cabo son los algoritmos a utilizarse.

4. Análisis de los Resultados. Verifica si los resultados obtenidos son coherentes con los obtenidos por el análisis y la visualización gráfica. Y el cliente determina si le aporta nuevos conocimientos que le permita la toma de decisiones.

Características de MD

- ✓ Explorar los datos que se encuentran en las profundidades de las bases de datos, o almacenes de datos, que algunas veces contienen información almacenada durante varios años.
- ✓ El entorno de la minería de datos suele tener una arquitectura cliente-servidor.
- ✓ Las herramientas de la minería de datos ayudan a extraer el mineral de la información enterrado en archivos corporativos o en registros públicos archivados.
- ✓ Las herramientas de la minería de datos se combinan fácilmente y pueden analizarse y procesarse rápidamente.
- ✓ La minería de datos produce cinco tipos de información:
    - Asociaciones.
    - Secuencias.
    - Clasificaciones.
    - Agrupamientos.
    - Pronósticos.

Aplicaciones de Minería de Datos

UNIDAD II:
Métodos para el Tratamiento y Análisis de Datos.

Introducción 1/2

La disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas ha transformado el análisis de datos orientándolo hacia determinadas técnicas especializadas englobadas bajo el nombre de minería de datos o Data Mining.

Introducción 2/2

Los métodos tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos.

Conceptos Básicos 1/2

- DATO:
  Un dato es un conjunto discreto de factores sobe un hecho real. Dentro de un contexto empresarial, el concepto de dato es definido como, un registro de transacciones.

Conceptos Básicos 2/2

- INFORMACIÓN:
  A diferencia de los datos la información tiene significado (relevancia y propósito). No solo pueden formar potencialmente al que la recibe, si no que esta organizada para algún propósito.

KDD
Extracción de información

KDD trata de interpretar grandes cantidades de datos para encontrar relaciones o patrones.

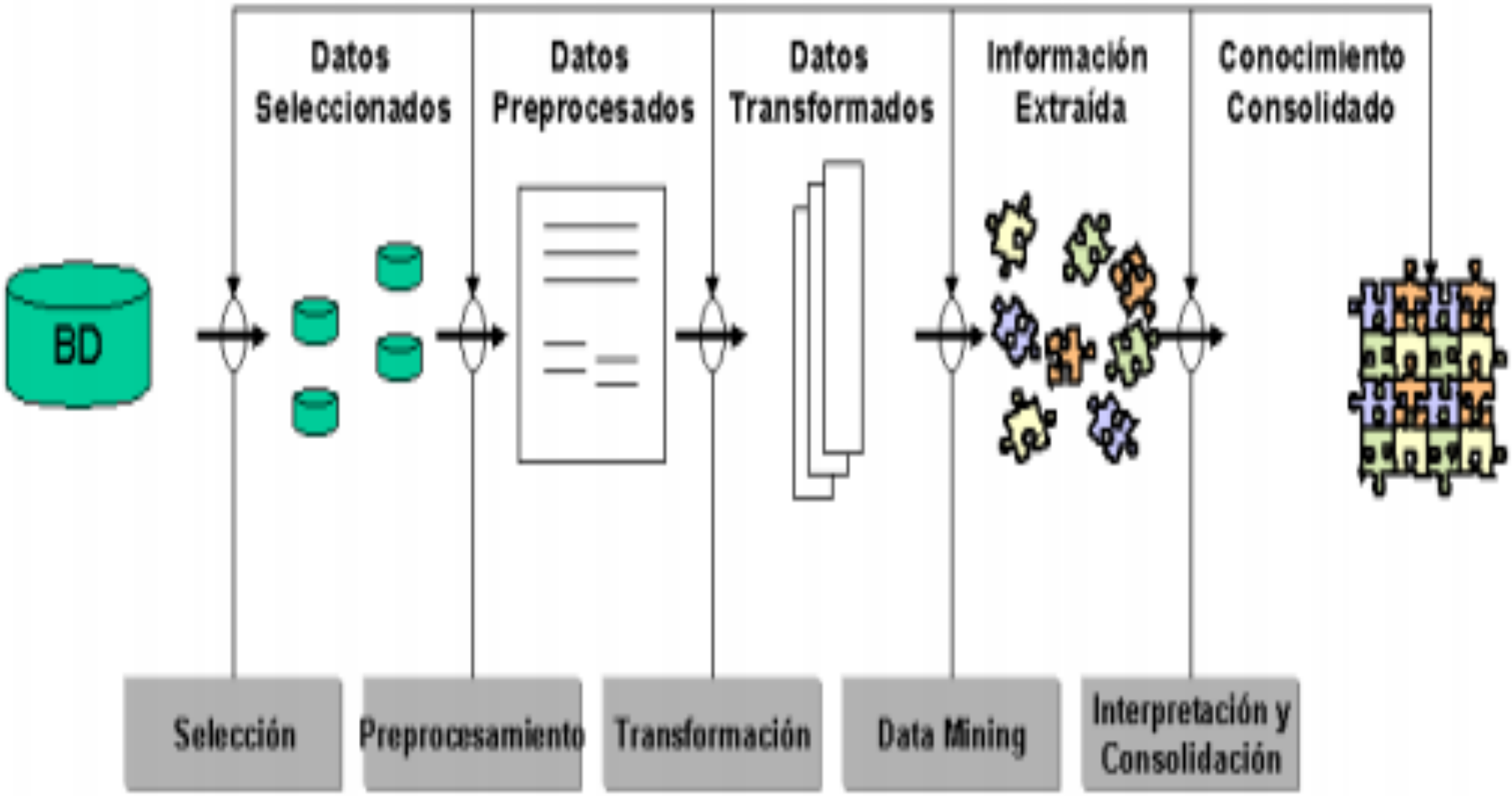
¿Cómo funciona?

1. Se inicia con la identificación de los datos
    - Qué datos se necesitan
    - Donde encontrarlos
    - Como conseguirlos

2. Seleccionar los datos útiles.

3. Seleccionar las herramientas y técnicas adecuadas.

Fases de la Minería de Datos 1/3



Fases de la Minería de Datos 2/3

Selección

- Recopilar e integrar las fuentes de datos existentes.
- Identificar y seleccionar las variables relevantes en los datos.
- Aplicar las técnicas de muestreo adecuadas.

Exploración

- Utilizar las técnicas de análisis exploratorio de datos.
- Deducir la distribución de los datos, simetría y normalidad.
- Analizar las correlaciones existentes en la información.

Fases de la Minería de Datos 3/3

Limpieza

- Detectar y tratar la presencia de valores inconsistentes.
- Imputar la información faltante o valores perdidos.
- Eliminar datos erróneos e irrelevantes.

Transformación

- Utilizar técnicas de reducción y aumento de la dimensión.
- Aplicar técnicas de numerización.

Redes neuronales artificiales:

Modelos predecible no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.

Arboles de Decisión:

Estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos.

Algoritmos genéticos:

Técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.

Método del Vecino Más Cercano:

Técnica que clasifica cada registro en un conjunto de datos basado en una combinación de las clases de los registros similares a él.

Regla de Inducción:

Extracción de reglas de datos basados en significado estadístico, para poder extraer o determinar la información importante en un volumen amplio de datos.

TÉCNICAS DE MINERÍA DE DATOS 6/6

Clustering (agrupamiento):

Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a todas las variables disponibles.

UNIDAD III:
Proceso de Análisis Supervisado.

Introducción 1/2

- Las técnicas de la minería de datos provienen de la inteligencia artificial y de la estadística. Dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener unos resultados. De entre las variadas técnicas, existen las llamadas reglas de asociación.
- Reglas de asociación: Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos. Según el objetivo del análisis de los datos, los algoritmos utilizados se clasifican en supervisados y no supervisados (Weiss y Indurkhya, 1998).

Introducción 2/2

- Algoritmos supervisados (o predictivos): predicen un dato (o un conjunto de ellos) desconocido a priori, a partir de otros conocidos.

- Algoritmos no supervisados (o del descubrimiento del conocimiento): se descubren patrones y tendencias en los datos.

Aprendizaje Supervisado 1/2

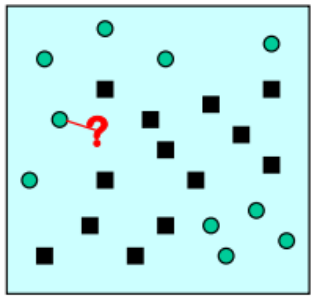
- Dependiendo de si se estima una función o una correspondencia:
  - Categorización: Se estima una correspondencia (las clases pueden solapar).
    - Ejemplo: determinar de un conjunto de temas de qué temas trata una determinada página web (cada página puede tratar de varios temas).
  - Clasificación: Se estima una función (las clases son disjuntas).
    - Ejemplo: determinar el grupo sanguíneo a partir de los grupos sanguíneos de los padres.
    - Ejemplo: Determinar si un compuesto químico es cancerígeno.

Aprendizaje Supervisado 2/2

- Técnicas:
  - k-NN (Nearest Neighbor).
  - k-means (competitive learning).
  - Perceptron Learning.
  - Multilayer ANN methods (e.g. backpropagation).
  - Radial Basis Functions.
  - Support Vector Machines
  - Decision Tree Learning (e.g. ID3, C4.5, CART).
  - Bayes Classifiers.
  - Center Splitting Methods.
  - Rules (CN2)
  - Pseudo-relational: Supercharging, Pick-and-Mix.
  - Relational: ILP, IFLP, SCIL.

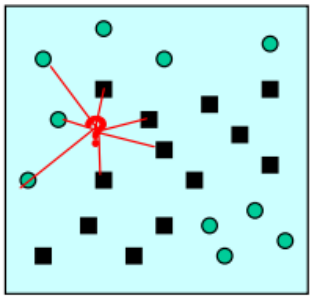
k-NN (Nearest Neighbour):

- 1. Se miran los k casos más cercanos.
- 2. Si todos son de la misma clase, el nuevo caso se clasifica en esa clase.
- 3. Si no, se calcula la distancia media por clase o se asigna a la clase con más elementos.



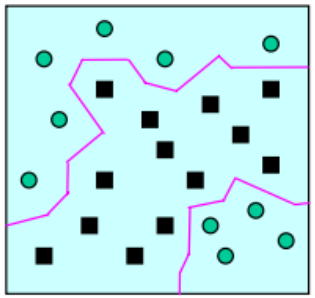
Clasifica círculo

1-nearest neighbor



Clasifica cuadrado

7-nearest neighbor



PARTICIÓN DEL 1-nearest neighbor (Poliédrica o de Voronoi)

Radial-Basis Function

- PRIMER PASO: Algoritmo Clustering:

  1. Dividir aleatoriamente los ejemplos en *k* conjuntos y calcular la media (el punto medio) de cada conjunto.

  2. Reasignar cada ejemplo al conjunto con punto medio más cercano.

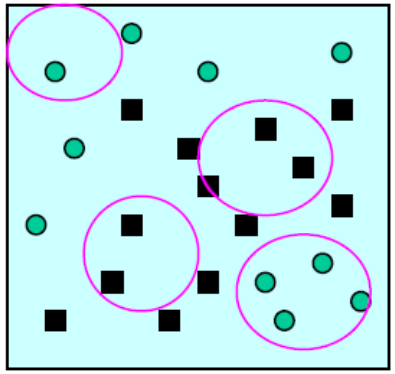
  3. Calcular los puntos medios de los *k* conjuntos.

  4. Repetir los pasos 2 y 3 hasta que los conjuntos no varíen.

- SEGUNDO PASO: Recodificar los ejemplos como distancias a los centros y normalizar.

Radial-Basis Function

- TERCER PASO: Con un perceptron de *k* elementos de entrada y una salida, aplicar el algoritmo visto antes.

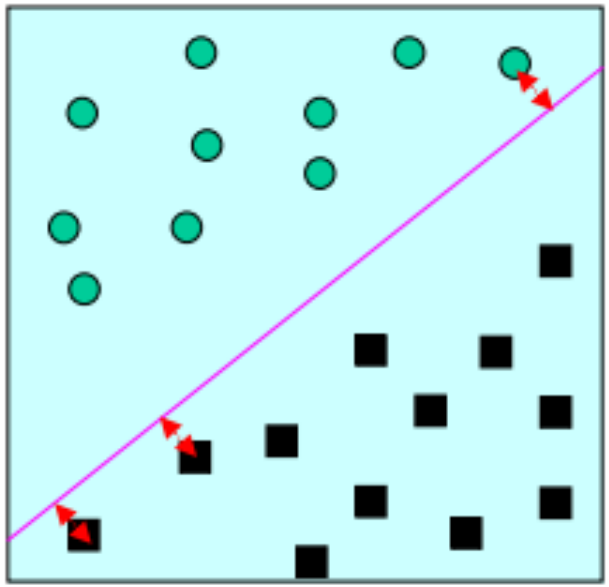


PARTICIÓN HIPERESFÉRICA CON 4 centros.

- Se convierte en una partición lineal (hiperplano) en un espacio de 4 dimensiones con los ejemplos siendo las distancias a los centros.

Máquinas de Vectores Soporte 1/2

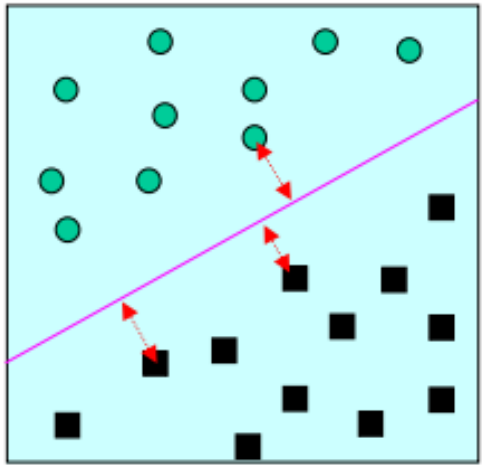
- Se basan en un clasificador lineal muy sencillo, precedido de una transformación de espacio (a través de un núcleo) para darle potencia expresiva.



Separa perfectamente, pero los tres ejemplos más cercanos (vectores soporte) están muy cerca de la frontera.

Máquinas de Vectores Soporte 2/2

- El clasificador lineal que se usa simplemente saca la línea (en más dimensiones, el hiperplano) que divida limpiamente las dos clases y además que los tres ejemplos más próximos a la frontera estén lo más distantes posibles.



Separa perfectamente, pero además los ejemplos más cercanos (vectores soporte) están lo más lejos posible de la frontera.

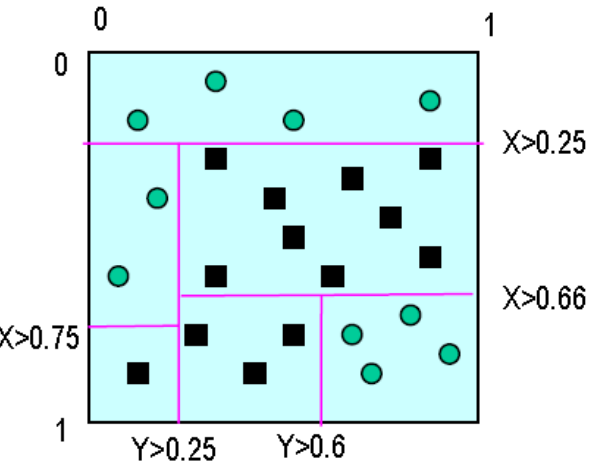
Árboles de Decisión (ID3 (Quinlan), C4.5 (Quinlan), CART) 1/4

Algoritmo Divide y Vencerás:

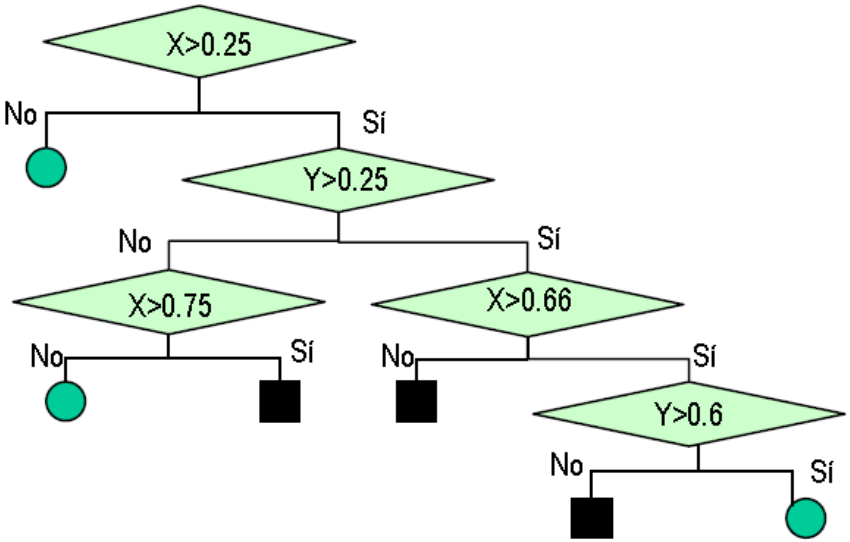
1. Se crea un nodo raíz con todos los ejemplos.
2. Si todos los elementos de S son de la misma clase, el subárbol se cierra. Solución encontrada.
3. Se elige una condición de partición siguiendo un criterio de partición (split criterion).
4. El problema queda subdivido en dos subárboles (los que cumplen la condición y los que no) y se vuelve a 2 para cada uno de los dos subárboles.

Árboles de Decisión (ID3 (Quinlan), C4.5 (Quinlan), CART) 2/4

- Algoritmo Divide y Vencerás:



PARTICIÓN CUADRICULAR.



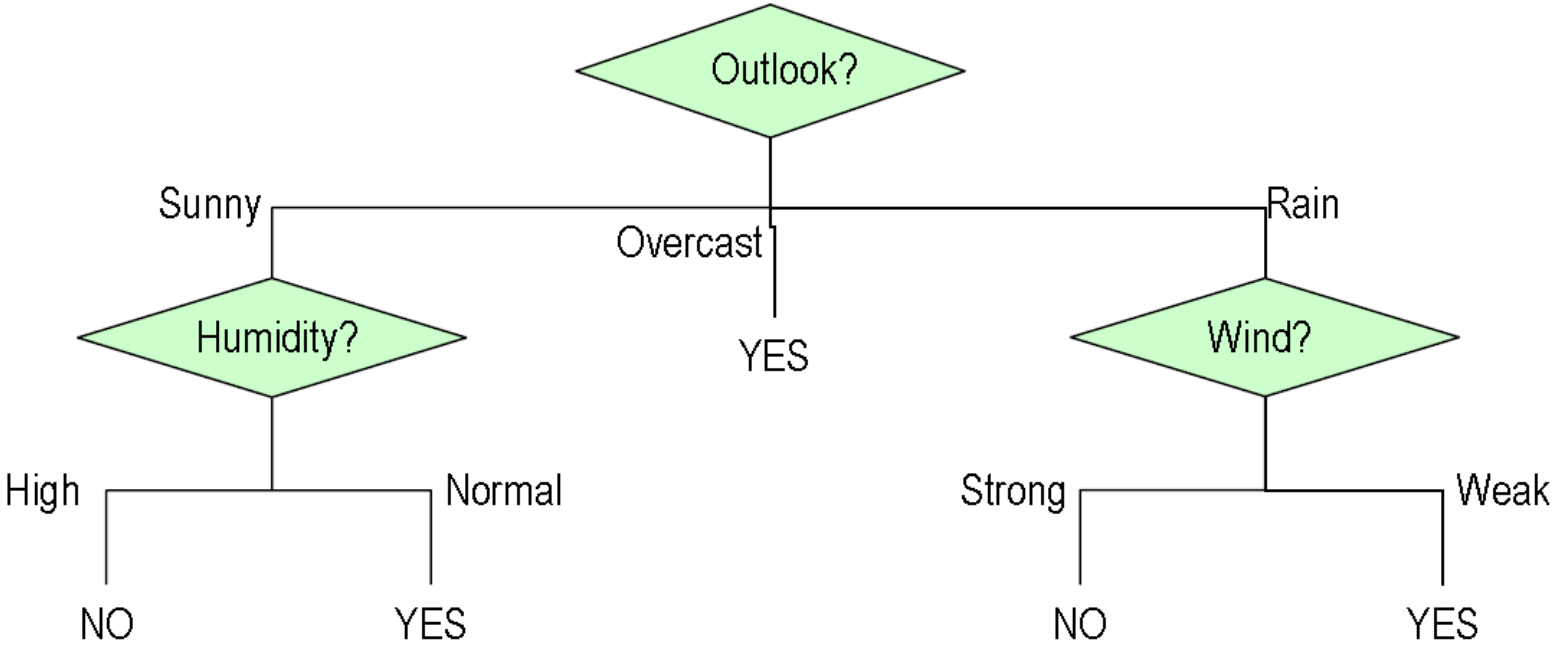
Árboles de Decisión (ID3 (Quinlan), C4.5 (Quinlan), CART) 3/4

Árboles de Decisión.

- **Ejemplo C4.5 con datos discretos:**

Example	Sky	Temperature	Humidity	Wind	PlayTennis
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cool	Normal	Weak	Yes
6	Rain	Cool	Normal	Strong	No
7	Overcast	Cool	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cool	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

Árboles de Decisión (ID3 (Quinlan), C4.5 (Quinlan), CART) 4/4



Representación Lógica:

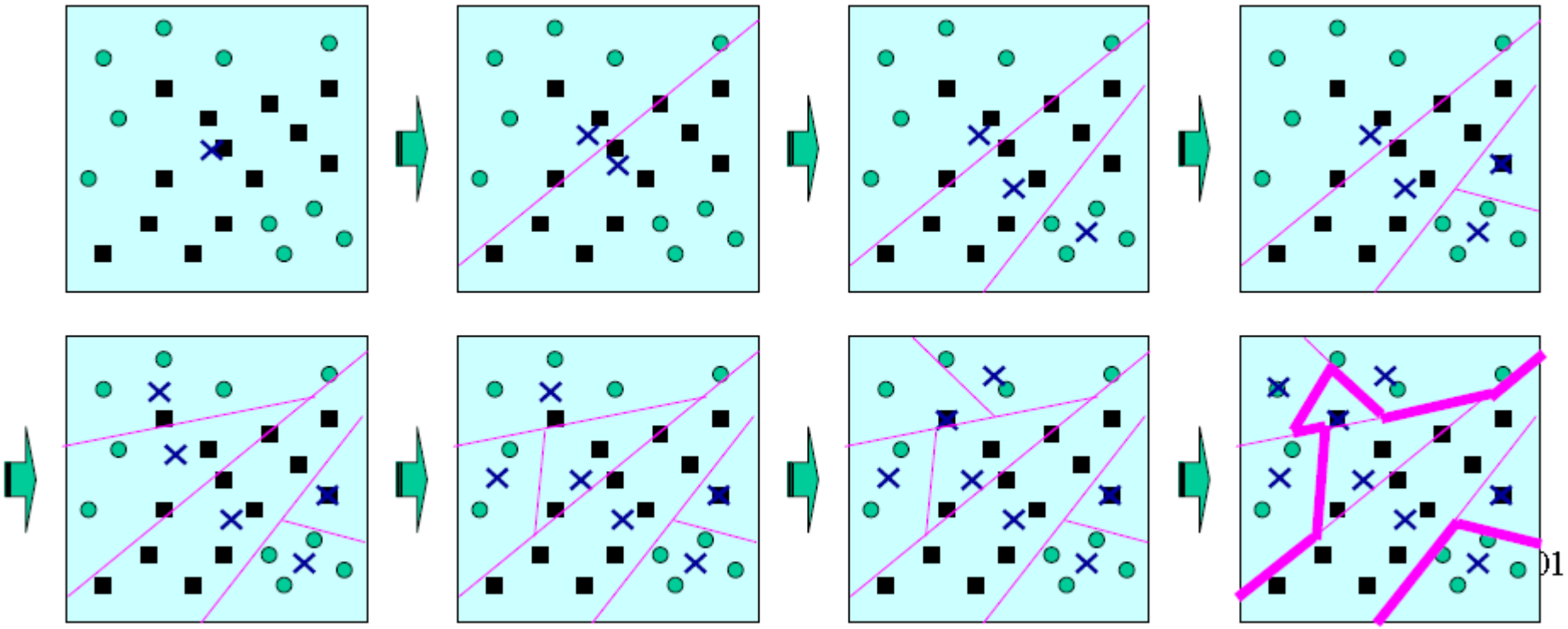
(Outlook=Sunny AND Humidity=Normal) OR (Outlook=Overcast) OR (Outlook=Rain AND Wind=Weak)

Center Splitting 1/2

Algoritmo:

1. Inicializar el primer centro en la media de los ejemplos.
2. Asignar todos los ejemplos a su centro más cercano.
3. Si hay algún centro que tiene ejemplos de diferente clase, borrar el centro y crear tantos nuevos centros distintos como clases haya, cada uno siendo la media de los ejemplos de la clase. Ir a 2.

Center Splitting 2/2



UNIDAD IV:
Proceso de Análisis No Supervisado.

Introducción 1/2

Los métodos no supervisados o también conocidos como el descubrimiento del conocimiento tiene como objetivo principal descubrir patrones, tendencias en los datos actuales y determinar que elementos ya sean genes o muestras presentan un patrón similar.

Introducción 2/2

La aplicación de los métodos no supervisados es descubrir los patrones de expresión que posteriormente podrán usarse en análisis supervisados.

Técnicas de los Métodos No Supervisados

Tenemos las Siguientes:

- Detección de desviaciones
- Segmentación
- Agrupamiento(“Clustering”)
- Reglas de asociación
- Patrones secuenciales
- Redes neuronales no supervisadas

Detección de Desviación

Frecuentemente estos objetos son conocidos como Outlier, la detección de anomalías también es conocida como detección de desviaciones , porque objetos anómalos tienen valores de atributos con una desviación significativa respecto a los valores típicos esperados.

Son tratados como ruido o error en muchas operaciones.

Agrupamiento o Clustering

El agrupamiento se puede considerar como la aproximación mas utilizada en aprendizaje no supervisado.

Su objetivo general es encontrar algún tipo de estructura en una colección de datos sin etiquetar o sin clasificar, ya que en la mayoría de los casos no se dispone de este tipo de información.

Redes Neuronales No Supervisadas

Estas redes son capaces de modificar sus parámetros internamente sin necesidad de supervisión.

Las redes neuronales no supervisadas por lo general tienen una arquitectura sencilla y se caracterizan por ser mas similares a los modelos biológicos que las redes neuronales supervisadas.

Reglas de Asociación

Las reglas de asociación en la minería de datos se utilizan para encontrar hechos que ocurren en común dentro de un conjunto de datos. Dicho de otra manera que debe ocurrir ciertas condiciones para que se produzca cierta condición.

Patrones Secuenciales

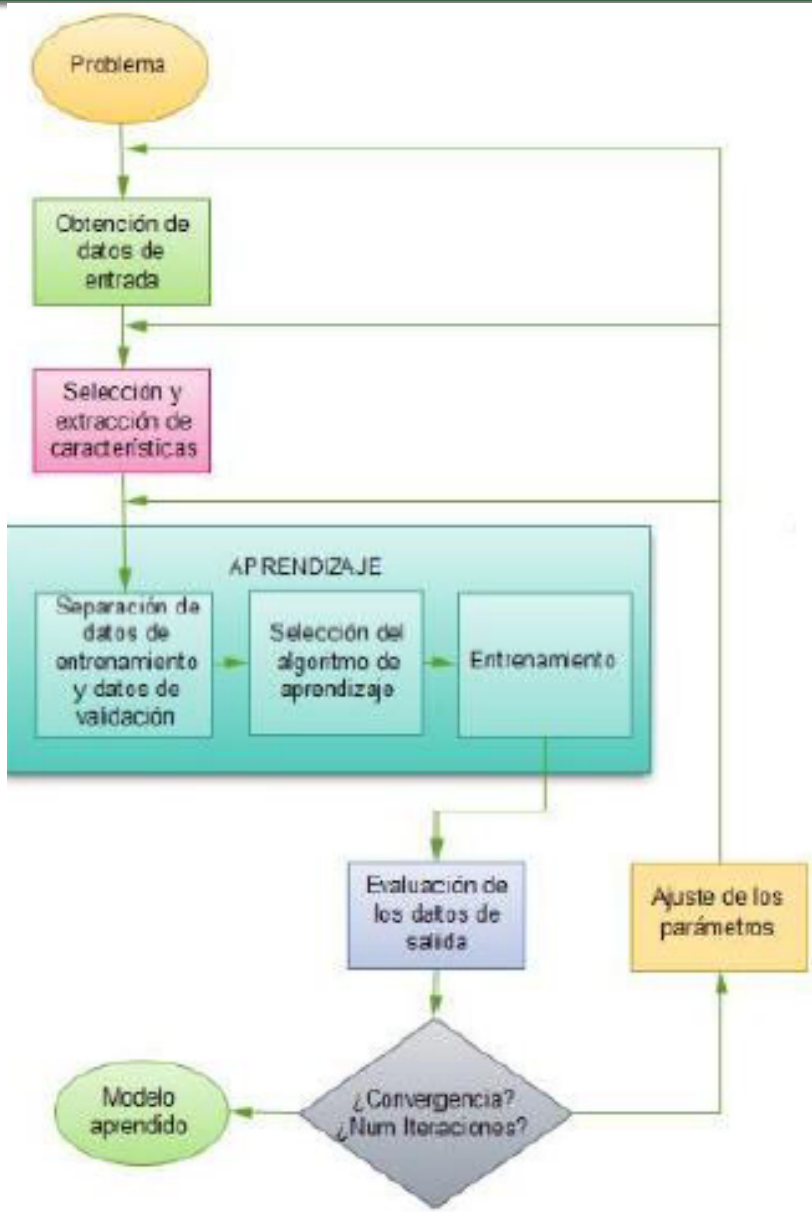
La minería de patrones secuenciales es la minería de patrones que ocurren frecuentemente relacionados al tiempo u a otras secuencias.

Aplicaciones de la minería de patrones secuenciales

Patrones de llamadas telefónicas, flujos de navegación en la web.

Estructuras de ADN y genes.

Proceso de un Aprendizaje No Supervisado



Fases del Proceso No Supervisado 1/2

- Las primeras fases son la obtención de datos y su preproceso (selección y extracción de características).
- En la fase selección y extracción de características el aprendizaje no es la misma, al no disponer de información acerca de la salida.
- En la fase de entrenamiento no se puede reajustar el modelo en base al error. Pero sigue siendo necesario separar los datos de entrenamiento y datos de validación para decidir si el método es bueno o no.

Fases del Proceso No Supervisado 2/2

- La fase de selección del algoritmo y el entrenamiento también se mantienen, en este caso la posibilidad de validar si los resultados son correctos no es frecuente, puesto que no se dispone de información de salida.
- La manera de decidir cuando se ha aprendido es viendo si el sistema converge o estableciendo un criterio de parada como puede ser un numero de iteraciones de funcionamiento máximo.

UNIDAD V:
Métodos Estimadores de Error.

Tareas de la Minería de Datos 1/2

El proceso de minería involucra ajustar modelos o determinar patrones a partir de datos. Este ajuste normalmente es de tipo estadístico, en el sentido que se permite un cierto ruido o error dentro del modelo.

Tareas de la Minería de Datos 2/2

A la hora de evaluar la capacidad predictiva de una herramienta de DM, el parámetro más importante suele ser la precisión de las predicciones que realiza. Para los sistemas de aprendizaje no supervisado, como análisis de conglomerados o generación de reglas de asociación.

Tasa de Error 1/2

La tasa de error es el complemento de la precisión, mide el porcentaje de las predicciones que son erróneas. Se suele utilizar cuando los niveles de precisión son muy altos, pues resulta más fácil apreciar la mejora.

Tasa de Error 2/2

Por ejemplo, la mejora de la precisión del 99,0% al 99,5% puede parecer menos importante que la mejora del 50% al 75%, sin embargo, en ambos casos la tasa de error se reduce a la mitad (una mejora espectacular).

Tasa de Error en Rechazo 1/3

A menudo, al realizar la predicción, el algoritmo de DM proporciona tanto la predicción como la confianza de que dicha predicción sea correcta.

Tasa de Error en Rechazo 2/3

Por ejemplo, el algoritmo del vecino más cercano puede proporcionar la misma predicción para todos los vecinos o para una mayoría. La predicción puede ser la misma en ambos casos, pero en el caso de unanimidad la confianza en la predicción es más alta.

Tasa de Error en Rechazo 3/3

Las predicciones pueden ordenarse según su confianza y las que menos confianza "inspiran" pueden rechazarse. De esta forma, se puede duplicar la precisión rechazando un 80% de predicciones.

Error Cuadrático Medio 1/2

Para las variables continuas, el grado de mal emparejamiento entre la predicción y el valor real pueden calcularse restando los dos valores y elevando el resultado al cuadrado. Este "error cuadrado" puede promediarse sobre todas las predicciones para estimar la distancia entre los valores reales y las predicciones.

Error Cuadrático Medio 1/2

La elevación al cuadrado tiene dos ventajas:

1. Por un lado, da un mayor peso a los errores graves.
2. Por otro lado, asegura que todos los errores son positivos y se suman a la hora de calcular la media.

Reduced-Error Pruning

Que consiste en dividir el conjunto de entrenamiento en n subconjuntos n-1 de los cuáles servirán realmente para el entrenamiento del sistema y 1 para la estimación del error. Sin embargo, el problema es que la construcción del clasificador se lleva a cabo con menos ejemplos.

UNIDAD VI:
Métodos para Análisis del Índice de Acierto.

Formas de Análisis

- Asumir distribuciones a priori.
- Criterio de simplicidad, de descripción o transmisión mínimas.
- Separar: Training Set y Test Set.
  - Cross-validation.
- Basadas en refuerzo.

Evaluación por Técnicas Bayesianas 1/2

- La mejor hipótesis es la más probable.
- Basadas en el teorema de Bayes. Despejan $P(h|D)$.
- La distribución de hipótesis a priori $P(h)$ y la probabilidad de unas observaciones respecto a cada hipótesis $P(D|h)$ deben ser conocidas.
- Son sólo técnicas evaluadoras aunque si el conjunto de hipótesis H es reducido se pueden utilizar en algoritmos de aprendizaje.

Evaluación por Técnicas Bayesianas 2/2

- Permiten acomodar hipótesis probabilísticas tales como “este paciente de neumonía tiene un 93% de posibilidades de recuperarse”.
- Muchas veces no se conoce P(h) o incluso P(D|h). Se hacen suposiciones: distribución uniforme, normal o universal.

Teorema de Bayes

Teorema de Bayes, y Maximum Likelihood (Probabilidad Máxima):

- $P(h|D)$: probabilidad de una hipótesis dado un conjunto de datos.
- $P(h)$: probabilidad a priori de las hipótesis.
- $P(D|h)$: probabilidad de D dada la hipótesis.
- $P(D)$: probabilidad a priori de los datos (sin otra información).

- Teorema de Bayes: (prob. a posteriori a partir de a priori)

$$P(h \mid D) = \frac{P(D \mid h)P(h)}{P(D)}$$

MAP (Maximum a Posteriori)

El Máximo a Posteriori se representa con la siguiente expresión:

- Criterio MAP (Maximum a Posteriori) (h es indep. de $P(D)$):

El Naive Bayes Classifier es un caso particular de este

$$h_{MAP} = \arg\max_{h \in H} P(h \mid D) = \arg\max_{h \in H} \frac{P(D \mid h)P(h)}{P(D)} = \arg\max_{h \in H} P(D \mid h)P(h)$$

Maximum Likelihood

El Máximo de Likelihood se representa con la siguiente expresión:

Maximum Likelihood (asumiendo $P(h)$ uniforme):

$$h_{ML} = \arg\max_{h \in H} P(D \mid h)$$

El Principio MDL (Minimum Description Length)

Asumimos $P(h)$ como la distribución universal (Occam's Razor):

$$P(h) = 2^{-K(h)}$$

donde $K(\cdot)$ es la complejidad descriptiva (Kolmogorov) de H.

FORMALIZACIÓN DE LA NAVAJA DE OCCAM:

"Las hipótesis con mínima descripción más pequeña son más probables".

Partición de la Muestra

- Evaluar una hipótesis sobre los mismos datos que han servido para generarla da siempre resultados muy optimistas.

  Solución: PARTIR EN: Training Set y Test Set.

- Si los datos disponibles son grandes (o ilimitados) :
  - *Training Set*: conjunto con el que el algoritmo aprende una o más hipótesis.
  - *Test Set*: conjunto con el que se selecciona la mejor de las anteriores y se estima su validez.
- Para problemas con *clase discreta*, se calcula la "accuracy", que se mide como el porcentaje de aciertos sobre el test set.
- Para problemas con *clase continua*, se utiliza la media del error cuadrático u otras medidas sobre el test set.

Accuracy

- Suponiendo la muestra *S* de *n* ejemplos, la hipótesis *h* es discreta y son independientes.
- Si n >= 30, nos permite aproximar la distribución binomial con la normal.
- Calculado el $error_s(h)$ sobre la muestra como *nºerrores/n*

Podemos obtener un intervalo de confianza a un nivel *c*:

$$error_S(h) \pm Z_c \cdot \sqrt{\frac{error_S(h)(1-error_S(h))}{n}}$$

Algunos valores de la tabla normal:

Nivel de confianza c:	50	68	80	90	95	98	99
Constante Zc:	0.67	1.00	1.28	1.64	1.96	2.33	2.58

Ejemplo Práctico

Considerando que una hipótesis da 12 errores sobre 40 ejemplos, por tanto, que con confianza 95% (Z_c = 1.96):

$$error_S(h) = 12/40$$

$$error_S(h) = 0.3$$

$$error_S(h) \pm Z_i \sqrt{error_S(h)(1-error_S(h))/n} = 0.3 \pm 1.96\sqrt{0.3(1-0.3)/40} = 0.3 \pm 0.14$$

Por lo tanto, para conseguir un nivel de confianza del 95%, es necesario que:

$$error_S(h) = 0.3 \pm 0.14$$

como $0.3+0.14=0.44$ *y* $0.3-0.14=0.16$, *entonces*,

el intervalo de errores permitido está entre 17.6 *y* 6.4

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Sumathi S., Sivanandam S.N. (2006). Introduction to data mining and its applications. Springer.
- Verma B., Blumenstein M. (2008). Pattern Recognition Technologies and Applications: Recent Advances. IGI Global.
- Eldén L. (2007). Matrix Methods in Data Mining and Pattern Recognition (Fundamentals of Algorithms). Society for Industrial and Applied Mathematics.
- Skillicorn D. (2007). Understanding Complex Datasets: Data Mining with Matrix Decompositions. Chapman & Hall/CRC Press.
- Wu X., Kumar V. (2009). The top ten algorithms in data mining. CRC Press.
- Tan P., Steinbach M., Kumar V. (2006). Introduction to Data Mining. Pearson Addison Wesley.

REFERENCIAS WEB

- http://www.it.uc3m.es/jvillena/irc/practicas/03-04/20.pres.pdf
- http://users.dsic.upv.es/~jorallo/cursoDWDM/dwdm-III-1.ppt
- http://exa.unne.edu.ar/informatica/SO/IM_2006.pdf
- http://inacap.serveftp.com/tic2/2_Prueba/02102014/mineria%20de%20datos.ppt
- http://adimen.si.ehu.es~rigauteachingEHUABDCurs%202005-2006EntregesBD%20emergetsPresentacion%20Data%20Mining.ppt

* De las cuales se tomaron imágenes para ilustrar este material.

UAEM