Summary Of Data mining

Radwan Mohammed

10/7/2014

Unit one	الوحده الاولى
Define Data?	**عرف البيانات ؟**
Data are any facts, numbers or text that can be processed by a computer.	هي حقائق أو ارقام أو نصوص تم معالجتها بإستخدام الحاسوب.
What are the types of data?	**عدد أنواع البيانات ؟**
• Operational Data • Non-Operational Data • Meta Data	• البيانات التشغيليه أو المتغيره • البيانات غير العمليه (مستقره) • البيانات نفسها
What is the different between Relational and Multidimensional database structure?	**ماهو الفرق بين مخطط قواعد البيانات العلائقي و المخطط متعدد الاتجاهات ؟**
• In a relational structure data is stored in tables permitting ad hoc queries. • In a multidimensional structure on other hands set of cubes are arranged in arrays with subset created according to category.	• في مخطط العلاقات يتم تخزين البيانات في جداول و السماح بإستعمال الاستعلامات للوصول اليها . • اما في متعدده الاتجاهات تكون البيانات عباره عن مجموعه من المكعبات و التي يتم ترتيب البيانات فيها بشكل مصفوفات و إنشاء مجموعه فرعيه منها.
What are things that can provide us information?	**ماهي الاشياء التي يمكن ان تزودنا بالمعلومات ؟**
• Patterns • Associations • Relationships	• الانماط • المجموعات • العلاقات
Note: information can be converted into knowledge about historical patterns and future trends.	ملاحظه: المعلومات يمكن أن تتحول الى معرفه حول انماط تاريخيه معينه أو بإتجاه المستقبل
Define Data mining?	**عرف تنقيب البيانات ؟**
Is a process of extracting hidden patterns from data.	هي عمليه إنتزاع الانماط المخفيه من البيانات
Explain The important of Data mining?	**أشرح اهميه تنقيب البيانات ؟**
• Data mining an increasingly important tool to transform this data into information. • Used in wide range of application such as marketing and fraud and scientific discovery.	• تنقيب البيانات تعتبر من اهم الادوات التي تستخدم في تحويل البيانات الى معلومات • تستخدم بشكل واسع في التطبيقات مثل التسوق و اكتشاف الخدع و الاكتشافات العلميه

Define knowledge discovery (Data m)?	**عرف إكتشاف المعرفه (أو تنقيب البيانات) ؟**
Is the process of analyzing data from different perspectives and summarizing it into useful information.	هي عمليه تحليل البيانات من أوجه مختلفه و تحليلها و تلخيصها لتصبح معلومات مفيده
Define data warehouse?	**عرف مستودع البيانات ؟**
Is a process of centralized data management and retrieval.	هو عمليه إداره و إسترجاع البيانات المركزيه
Note: centralization of data is needed to maximize user access and analysis.	ملاحظه : البيانات المركزه تحتاج الى مستخدم كحد اقصى للوصول للبيانات و تحليلها .
What are Data mining tasks?	**عدد مهام تنقيب البيانات ؟**
• Classification • Clustering • Association • Regression	• التصنيف • المجموعات • الروابط • الانحدار
Define the classification?	**عرف التصنيف؟**
Arranges the data into predefined groups for example the Email Working with 2 algorithms Nearest neighbor and Neural network	هو عمليه ترتيب البيانات داخل مجموعات معرفه مثل الايميل . وتعمل بخوارزميتين اقرب جار و الشبكه العصبيه
Define clustering?	**عرف المجموعات ؟**
Give a set of data point each having a set of attributes and similarity measure Data in one cluster are more similar to one another	تعطي مجموعه من النقاط من البيانات وكل مجموعه تمتلك مجموعه من الخصائص و لها نفس المقياس كلما كانت نقطه البيانات في نفس المجموعه كان التشابه كبير
Define Association Rule Discovery?	**عرف إكتشاف قاعده الربط ؟**
Given a set of records each of which contain some number of items from a given collection. Searches for relationships between variables.	يعطي مجموعه من الحقول وكل حقل يحتوي على مجموعه من العناصر نلاحظ أكثر العناصر إرتباطا مع العناصر الاخرى و نعتبرها هي قاعده الارتباط أو هي البحث بين المتغيرات عن علاقه تربطهم
Note: the clustering is like classification but the groups are not predefined	ملاحظه: المجموعات تشبه التصنيفات لكن المجموعات لا تكون معرفه

Define Regression?	**عرف الانحدار؟**
Attempts to find a function which models the data with least error And used Genetic programming	هو محاوله أيجاد داله لتشكيل وتمثيل البيانات مع اقل عدد ممكن من الاخطاء بإستخدام البرمجه الوراثيه
What are data mining elements?	**عدد عناصر تنقيب البيانات ؟**
• Extract, transform, and load transaction data onto the data warehouse system. • Store and manage the data in a multidimensional database system. • Provide data access to business analysts and information technology professionals. • Analyse the data by application software. • Present the data in a table.	• تخزين و نقل البيانات و تحميل البيانات الى نظام تخزين قواعد البيانات • تخزين و إداره البيانات بإستخدام أنظمه قواعد البيانات المتعدده • تمكين البيانات من الوصول الى تحليل العمليات و المعلومات الاحترافيه • تحليل البيانات بواسطه التطبيقات • تمثيل البيانات و تخزينها في جداول
What are analysis levels ?	**عدد مستويات التحليل ؟**

• **Artificial neural networks**: Non-linear predictive models that learn through training and resemble (imitate) biological neural networks in structure.
• **Genetic algorithms**: Optimization techniques that use processes such as genetic combination, mutation (change), and natural selection in a design based on the concepts of natural evolution.
• **Decision trees**:
  o Tree-shaped structures that represent sets of decisions.
  o These decisions generate rules for the classification of a data set.
• **Rule induction**: The extraction of useful if-then rules from data based on statistical significance.
• **Data visualization**: The visual interpretation of complex relationships in multidimensional data.
	• الشبكه العصبيه الاصطناعيه: هي عباره عن نموذج غير خطي يتعلم بالتدريب وهو يشبه الجهاز العصبي الطبيعي
	• الخوارزميه الجينيه : هي عباره عن تقنيات إختياريه تستخدم مجموعه وراثيه لتصميم المفاهيم الاساسيه في الوراثه الطبيعيه
	• شجره القرار: هي تمثل نفس هيكل الشجره الطبيعي وهي مجموعه من القرارات التي تولد قواعد لتصنيف البيانات
	• قاعده زياده الانتاجيه : هي عمليه إستخلاص القواعد المهمه (إذا كان فإن) من البيانات الاساسيه في المستندات الاحصائيه
	• البيانات الافتراضيه : هي تمثيل العلاقات المعقده في بيانات متعدده الابعاد

Note: data mining applications are available on all size systems for mainframe, client/server, PC platforms	ملاحظه : تطبيقات تنقيب البيانات متاحه في كل الانظمه client/server , PC , mainframe
Data mining do tow processes what are this?	**يقوم تنقيب البيانات بعمليتين رئيسيتين إذكرهما ؟**
• Discovery • Prediction	• الاستكشاف • التنبئ
What are the applications that uses in Data mining ?	**عدد التطبيقات المستخدمه في تنقيب البيانات؟**
• RapidMiner • Weka • Art	• RapidMiner • Weka • Art
What are the data mining issues?	**ماهي قضايا تنقيب البيانات ؟**
• **1. Business issues**: analysing routine business transactions and classifications. • **2. social issues**: • **3. Mining Methodology Issues:** Pertain to data mining approaches applied and their limitations. • **4. Cost:** While system hardware costs have dropped dramatically within the past few years, data mining and data warehousing tend to be self-reinforcing • **5. User Interface Issues:** The knowledge discovered by data mining tools is useful as long as it is interesting, and above all understandable by the user. • **6. Data Source issue:** An excess of data appear when we have more data than we can handle - different types of data are stored in a variety of repositories	• القضايا التجاريه: هي عمليه تحليل البيانات التجاريه الموجهه و تحويلها و تصنيفها . • القضايا الاجتماعيه • قضايا منهجيه التنقيب : مناسبه وملائمه لتنقيب البيانات و يطبق في تحديدها • التكلفه : عند حدوث توقف دراماتيكي لتكلفه في السنوات الماضيه فإن تنقيب البيانات و تخزينها يوفر الدعم الذاتي لحل هذه المشكله • قضايا واجهات المستخدم : المعرفه المكتشفه بواسطه تنقيب البيانات تكون مفيده ومفهومه للمستخدم وهو يحتاج الى نتيجه جيده وتناظريه لتنقيب البيانات • قضايا مصادر البيانات :هي مجموعه من البيانات المنفذه و التي تظهر عندما نمتلك الكثير من البيانات التي يمكن ان نستعملها و قد تكون مجموعه من الانواع المخزنه في تشكيله من الانماط

What is Data mining software?	عرف برامج تنقيب البيانات ؟
Data mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, categorize it, and summarize the relationships identified.	هي واحده من الادوات المستعمله في تحليل البيانات بحيث يسمح للمستخدم بتحليل البيانات من عدد كبير من المصادر و تلخيصها و تحديد علاقات الربط بينها .
What is Technically of Data mining?	**عرف تقنيه تنقيب البيانات ؟**
Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. And that are two groups data mining tools and data mining applications	هي عمليه إيجاد إرتباطات و انماط من بين عشرات الحقول في قواعد البيانات العلائقيه الكبيره ويتم تقسيمها الى 2 : ادوات تنقيب البيانات وتطبيقات تنقيب البيانات
Note: Organizations are using data mining tools and data mining applications together in an integrated environment for predictive analytics.	ملاحظه : لعمل إنتاج تحليلي مميز يتم إستعمالهما معاً الادوات و التطبيقات و كلاهما متاح
What are the goals of Data mining tools	**عدد المهام التي تقوم بها ادوات تنقيب البيانات ؟**
Data mining tools provide both developers and business users with an interface for discovering, manipulating, and analysing corporate data	توفر للمستخدم الشاشات التي تساعده على - الاكتشاف - المعالجه - التحليل
Explain Text mining and web mining?	**اشرح تنقيب النصوص و تنقيب الانترنت ؟**
Recent advances have led to the newest and hottest trends in data mining—text mining and Web mining. These two data mining technologies open a rich vein of customer data in the form of textual comments from survey research and log files from Web servers	فوائد جديده يمكن ان تقود الى مجموعه جديده وموجهه من تنقيب البيانات وهذه التقنيات تغذي العميل بالبيانات حيث تسمح له بإستعراضها خلال تصفح الانترنت و تحليلها وإضافه المعلومات الى السرفر

Unite TOW	الوحده الثانيه
Note: data mining is the core of KDD	ملاحظه: تنقيب البيانات تعتبر نواه الـ KDD
Define KDD (knowledge Discovery in Database)?	**ماهو تعريف KDD عمليه اكتشاف المعرفه في قواعد البيانات ؟**
process of finding useful information and patterns in data.	هي عمليه ايجاد المعلومات و الانماط المفيده من البيانات
What are data mining algorithm components?	**ماهي مكونات خوارزميات تنقيب البيانات ؟**
• Model representation descriptions of discovered patterns • Model evaluation criteria how well a pattern (model) meets goals • Search method parameter search: optimization of parameters for a given model representation	• نموذج التمثيل : يقوم بوصف الانماط المكتشفه • نموذج تقدير المقياس : وصف كيف يمكن لهذه الانماط تحقيق الاهداف • طريقه البحث : • معامل البحث ينظم الانماط لاعطاء نموذج تماثلي
Note: Data mining involves fitting models to and determining patterns from observed data	ملاحظه : تنقيب البيانات يشمل تركيب النماذج و تحديد الانماط من مجموعه من النماذج المعروضه
What are the steps involved in KDD process?	**ماهي الخطوات التي تتم في معالجه ال KDD ؟**
• ***Selection:*** Obtain data from various sources. • ***Preprocessing:*** data cleaning. • ***Transformation:*** Convert to common format. Transform to new format. • ***Data Mining:*** Obtain desired results by applying Data Mining tasks tools. • ***Interpretation/Evaluation:*** Present results to user in meaningful manner.	• الاختيار : أختيار البيانات من مصادر مختلفه • الاعداد : تنظيف البيانات • النقل : نقل البيانت الى إطار جديد • تنقيب البيانات : أختيار النتيجه المطلوبه من خلال إستخدام أدوات ومهام تنقيب البيانات • التفسير : عرض و تمثيل النتائج للمستخدم .

What are the stages of data mining process?	**ماهي المراحل التي تتم في عمليه تنقيب البيانات؟**
Consists of three stages: (1) The initial exploration, (2) Model building (3) Deployment	عمليات تنقيب البيانات تتكون من 3 مراحل: 1- الاستكشاف الداخلي 2- بناء النموذج 3- الانتشار
Explain Exploration? (stage one)	**اشرح مفهوم الاكتشاف الاولي؟**
This stage usually starts with data preparation which may involve cleaning data, data transformations, selecting subsets of records.	في هذه المرحله نبداء بإعداد البيانات و التي تشمل تنظيف البيانات و تحويلها وإختيار الحقول الفرعيه
Where EDA used?	**في ماذا تستخدم تقنيه EDA ؟**
Exploratory Data Analysis (EDA) is used to identify systematic relations between variables when there are no (or not complete) expectations as to the nature of those relations.	تستخدم في توضيح العلاقات المنظمه بين المتغيرات عندما لاتكون مكتمله فيقوم بتوقع العلاقات الطبيعيه فيها
Explain Model Building? (stage tow)	**اشرح معنى بناء النموذج؟**
choose the suitable models to represent the explored data	إختيار النموذج الملائم لتمثيل البيانات المكتشفه
Explain Deployment? (stage three)	**اشرح معنى الانتشار ؟**
in deployment ensure that the resultant patterns meet the required patterns for prediction and decision making	هو التأكد بان الانماط الناتجه قابلت الانماط المطلوبه للتنبئ و إتخاذ القرار
What are data mining functionalities?	**ماهي وظائف تنقيب البيانات ؟**
• -Characterization: summarization of general features of objects and produces characteristics rules. • - Discrimination: Comparison between two classes, *target class* and *contrasting* class • - Association analysis: the frequency of items occurring together in transactional database. • - Classification: Organization of data in a given class.	• الوصف : هو ملخص للمميزات العامه للعناصر و إنتاج القواعد المميزه • الاختلاف: مقارنه بين نوعين من التصنيف (تصنيف الهدف و التصنيف المناقض) • تحليل الروابط : هو عمليله تكرار العناصر التي تحدث معا في عمليه نقل قواعد البيانات • التصنيف : هو تنظيم البيانات في أصناف معطاه

What are the types of prediction?	ماهي أنواع التنبئ ؟
• predict some unavailable data values • predict a class label for some data	• التنبئ للقيم الغير متاحه • التنبئ بالتصنيف لبعض البيانات
What is Outlier analysis?	**ماهو التحليل الخارجي؟**
Outliers are data elements that cannot be grouped in a given class or cluster. Known as exceptions or surprises. In some applications they are noise, but they can reveal important knowledge in other domains.	هي عباره عن بيانات لا تستطيع أن تكون مجموعه في تصنيف معين أو تجمع (تعرف بالاستثنائات) في بعض التطبيقات تعتبر ضوضاء لكن يمكن ان تكون مهمه في بعض التطبيقات الاخرى
What is the different between Evolution and deviation analysis?	**ماهو الفرق بين الانحراف المعياري و التدرج ؟**
• Evolution pertain to the study of time related data that changes in time. • Deviation analysis considers differences between measured values and expected values.	• الانحراف المعياري: هو دراسه التغيرات التي تحصل في التحليل خلال فتره زمنيه معينه. • التدرج: هو الفرق بين القيمه الفعليه و القيمه المتوقعه.
Unite three	**الوحده الثالثه**
What are data processes?	**ماهي عمليات البيانات ؟**
• Data Cleaning • Data Integration • Data Transformation • Data Reduction	• تنظيف البيانات • تكامل البيانات • تحويل البيانات • فصل البيانات أو إختصارها
What are data cleaning capabilities include?	**ماهي العمليه التي تتم عند تنظيف البيانات؟**
• Smoothing noisy data • -Eliminate duplicate records • -Identification of missing or incomplete data • -Removal of obsolete (not used) data	• تنعيم البيانات المزعجه • التخلص من الحقول المكرره • تعيين البيانات المفقوده و غير المكتمله • حذف و إزاله البيانات المهمله او غير المستعمله
What is noise data?	**ماهي البيانات المزعجه ؟**
Noise is a random error or variance in a measured or recorded data	هي أخطاء عشوائيه أو إختلاف في المقياس او في حقول البيانات

How can we smoothing data in DM?	**كيف يمكن تنعيم البيانات في تنقيب البيانات ؟**
In data mining binning method is used to smooth data	يتم إستعمال الصناديق لتقسيم و تصنيف البيانات ومن ثم تنعيمها
Given a numerical attribute such as *Price* with data: 3,27,7,32,25,25,6,28,22	
Using Binning (with three bins) will give • Partitioning Bin 1: 3 6 7 Bin 2: 22 25 25 Bin 3: 27 28 32 • Smoothing by Bin Mean (for the nearest recorded value) Bin 1: 6 6 6 Bin 2: 25 25 25 Bin 3: 28 28 28	
Suppose a group of 12 sales price records has been stored as following: 5,10,11,13,15,35,50,55,72,92,204,215 **Partition them into three bins by each of the following methods :** **a- Equal frequency partitioning** **b- Equal width partitioning** **c- Clustering**	
a- Bin 1: 5 10 11 13 Bin 2: 15 35 50 55 Bin 3: 72 92 204 215 b- ? c- Clustering A={ 5,10,15,35,50,55,215} B={11,13,72,92,204}	
What is data integration?	**ماهو تكامل البيانات ؟**
Combining data from multiple data stores into a coherent data store as in data warehousing.	خلط البيانات من مخازن بيانات متنوعه و البيانات المتماسكه تم تخزينها في مستودع البيانات
What are data transformation processes?	**ماهي عمليات تحويل البيانات ؟**
Aggregation,Generalization Normalization,Feature Construction	الاجمالي و عباره عامه و التطبيع و تقييم البناء
What is the meaning of normalization?	**ماهو تطبيع البيانات ؟**
In Normalization attribute data are scaled so as to fall within small specified range. Useful for classification and clustering.	هي عمليه إختصار الخصائص الى مدى صغير ومحدود وهو مفيد لعمليه التصنيف والمجموعات

What are normalization techniques?	ماهي تقنيات و انواع تطبيع البيانات ؟
• *Min-Max Normalization* • *Z-Score normalization*	• التطبيع الاكبر و الاصغر • التطبيع النهائي

Min-Max Normalization:

$$ú = \frac{v - min_A}{max_A - min_A} \times (new_max_A - new_min_A)\ new_min_A$$

Z-Score normalization:

$$ú = \frac{v - Ã}{\sigma_A}$$

where:

Ã is the mean value

σ_A is the standard deviation

Consider min and max values for the attribute *income* are $12,000 and $98,000. Map range = [0.0, 1.0] or $min_A = 0$, $max_A = 1.0$ then a value of *v*=$73.600 for income is transformed to: What is the value of normalization?

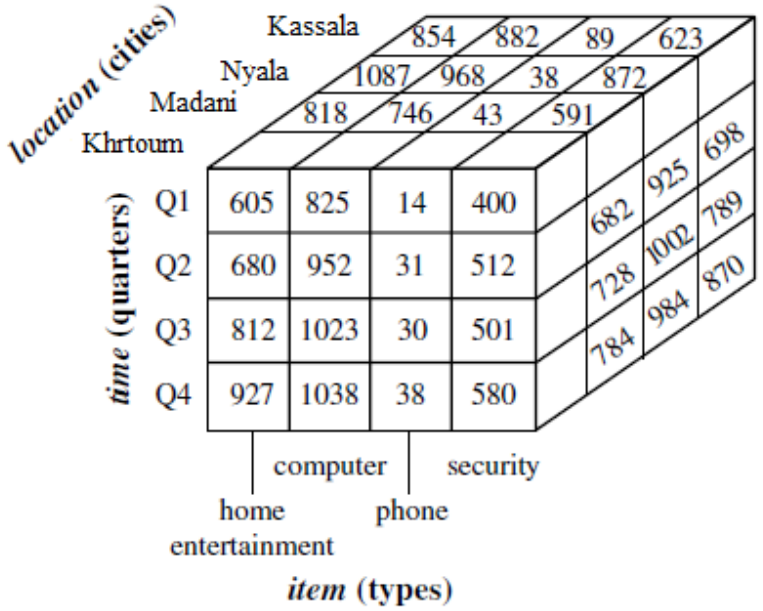
$$\frac{73{,}600 - 12{,}000}{98{,}000 - 12{,}000} \times (1.0 - 0.0) + 0 = 0.716$$

Consider the mean and standard deviation of the values for the attribute income are $54,000 and $16,000 respectively, with z-score normalization, a value of $73,600 for income is transformed to:

$$\frac{73{,}600 - 54{,}000}{16{,}000} = 1.225$$

Explain Data Reduction?	أشرح عمليه فصل البيانات ؟
Data mining on huge amounts of data is impractical and takes a long time. Data reduction is useful for obtaining reduced data set without losing its integrity.	تنقيب البيانات في محتوى ضخم من البيانات غير عملي ويأخذ الكثير من الوقت . فصل البيانات مفيد لكسب مجموعه من البيانات بدون حدوث فقد في البيانات و جعلها متكامله
There are some steps for reduction data?	**هناك عده مراحل لفصل البيانات ماهي ؟**
Data cube aggregation, Attribute subset selection, Histograms	البيانات المجموعه المكعبه – إختيار الخصائص الفرعيه و المنحنى التكراري

Draw a 3-D data cube representation of the data in Table below according to time , Time ,Item , and location (Khartoum , Nyala , Kassala , Medani)
The answer :



Unite Four	الوحده الرابعه
What are data mining techniques?	**ماهي تقنيات تنقيب البيانات ؟**
• Classification • Decision Tree • Neural Networks • Genetic Algorithms	• التصنيفات • شجره القرارات • الشبكه الصناعيه • الخوارزميه الجينيه
Note: Prediction predicts unknown or missing values.	ملاحظه : التنبئ قد يتنبئ قيم غير معروفه او قيم مفقوده.
What is decision tree and what are his parts?	**ماهي شجره القرارات وماهي اجزائها ؟**
is a computational model consisting of three parts: • Decision Tree • Algorithm to create the tree • Algorithm that applies the tree to data	هي عباره عن نموذج حسابي يتكون من 3 اجزاء : • شجره القرار • الخوارزميه لانشاء الشجره • الخوارزميه التي تطبق الشجره على البيانات

What are DT advantages/disadvantages?	ماهي مميزات و عيوب شجره القرارات ؟

• *Advantages:*
  ○ Easy to understand.
  ○ Easy to generate rules
• *Disadvantages:*
  ○ May suffer from overfitting.
  ○ Classifies by rectangular partitioning.
  ○ Does not easily handle nonnumeric data.
  ○ Can be quite large – pruning is necessary.
	• المميزات :
	  ○ سهل في الفهم
	  ○ سهل في توليد القواعد
	• عيوبها :
	  ○ يمكن ان تعاني من زياده المقياس
	  ○ التصنيف يكون بالمستطيلات المقسمه
	  ○ ليست سهله للبيانات غير العدديه
	  ○ يمكن ان تكون ضخمه في حاله ضروره التجزئه

What is neural network?	**ماهي الشبكه الاصطناعيه؟**
Is a collection of processing nodes transferring activity to each other via connections (the brain).	هي مجموعه من العقد المتحوله و النشطه و المترابطه مع بعضها البعض مثل المخ
Explain Artificial network?	**اشرح مفهوم الشبكه العصبيه؟**
In Artificial Neuron all signals can be 1 or -1 as a binary case often called classic spin. The neuron calculates a weighted sum (X) of the inputs, and compare it with a Threshold (T). If the input is higher than Threshold T, the output is set to 1, otherwise to -1. Output S either 1 or -1.	في الشبكه العصبيه تكون كل القيم و الاشارات بين 1، -1 نقوم بحساب مجموع الاوزان المدخله ومقارنتها مع قيمه العتبه T اذا كانت اكبر منها = 1 اذا كانت اصغر منها = -1 اذا كانت تساويها = 0
What is feed forward approach?	**ماهي التغذيه العكسيه ؟**
NN is trained to classify certain patterns into certain groups, and then used to classify new patterns presented to the net.	تقود انماط التصنيف الى مجموعات مركزيه و استخدامها في تصنيف انماط جديده و تمثيلها و عرضها في الانترنت
What are the components of Genetic Algorithm?	**ماهي مكونات الخوارزميه الجينيه ؟**
• Flags • Relation operator • Values	• الاعلام : له حالتان =1 اي اعرض الشرط الذي يتناسق مع الشرط =0 اعرض الشرط الذي سوف يحذف من القاعده الارتباط له حالتان اذا كانت صريحه =and= ومكمله =and>

Explain OLAP?	**اشرح مفهوم الـ OLAP ؟**
On Line Analytical Processing performs multidimensional analysis of business data and provide capability for sophisticated data modelling. ROLAP - Relational OLAP MOLAP - Multidimensional OLAP	هي تحليل علمي يمثل أبعاد متعدده من تحليل البيانات التجاريه و تزويدها بالكفائه و الخبره لها نوعين : الـ Relational OLAP - ROLAP الـ Multidimensional OLAP – MOLAP
Note: ***(OLAP):*** provides more complex queries than OLTP.	ملاحظه : الـ OLAP يزودنا بإستعلامات معقده أكثر من الـ OLTAP
What are OLAP operations?	**عدد عمليات الـ OLAP ؟**
• Single cell • Multiple cell • Slice • Dice	• خليه فرديه • خليه متعدده • شريحه • نرد
Unit Five	**الوحده الخامسه**
What is Estimation Error?	**ماهو توقع الخطاء؟**
Difference between expected value and actual value. $Bias = E(\hat{\Theta}) - \Theta$	هو الاختلاف او الفرق بين القيمه المتوقعه و القيمه الفعليه

MLE= $L(\Theta \mid x_1, ..., x_n) = \prod_{i=1}^{n} f(x_i \mid \Theta)$

Coin toss five times: {H,H,H,H,T}

Assuming a perfect coin with H and T equally likely, the likelihood of this sequence is:

$$L(p \mid 1, 1, 1, 1, 0) = \prod_{i=1}^{5} 0.5 = 0.03.$$

However if the probability of a H is 0.8 then:

$$L(p \mid 1, 1, 1, 1, 0) = 0.8 \times 0.8 \times 0.8 \times 0.8 \times 0.2 = 0.08.$$

Variance(التباين) & Standard Deviation (الانحراف المعياري)

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{\infty}(x_i - \bar{x})^2$$

التباين

$$\sigma = \sqrt{\frac{1}{N}\sum_{n=1}^{\infty}(x_i - \bar{x})^2}$$

الانحراف المعياري

Note: σ = 0 only when there is no spread	ملاحظه : تكون قيمه الانحراف المعياري = 0 عندما لا يكون هناك إنتشار
Explain Regression?	اشرح مفهوم الانحدار ؟
• The **unknown parameters** denoted as **β**. This may be a scalar or a vector of length *k*. • The **independent variables**, **X**. • The **dependent variable**, *Y*. Y = f (**X**, **β**)	يمكن ان يستخدم في التنبئ داخل سلسله زمنيه من البيانات × = مستقل ، Y = تابع ، B= عنصر غير معروف

قانون ايجاد القيمه المتوقعه

$$\chi^2 = \sum \frac{(O-E)^2}{E}.$$

O={50,93,67,78,87}

E=75

$$X^2 = \frac{(50-75)^2 + (93-75)^2 + (67-75)^2 + (78-75)^2 + (87-75)^2}{75 \quad 75 \quad 75 \quad 75 \quad 75} = 15.54$$

Examine the degree to which the values for two variables behave similarly. Correlation coefficient r: 1 = perfect correlation -1 = perfect but opposite correlation 0 = no correlation	الربط : يفحص درجه كل القيم لكل قيمتين يتصرفان بشكل متشابه لو = 1 يكون إرتباط مثالي لو = 1- يكون إرتباط مثالي عكسي لو = 0 لا يوجد هناك إرتباط

$$r = \frac{\sum(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum(x_i - \bar{X})^2 \sum(y_i - \bar{Y})^2}}$$

Good Luck

Radwan Mohammed

Aljaki2@live.com

7/10/2014 2:02 AM