Lecture – 23
Big Data Machine
Learning (Part-I)

Big data, Machine Learning.

Refer Slide Time :(0: 16)

Preface

Content of this Lecture:

- In this lecture, we will provide an overview of machine learning techniques to explore, analyze, and leverage data.

- We will also discuss tools and algorithms that you can use to create machine learning models that learn from data, and to scale those models up to big data problems.

Preface content of this lecture: In this lecture, we will provide an overview of machine learning techniques, to explore, analyze and leverage, big data processing, we will also, discuss the tools and algorithm that you can use, to create machine learning models and learn from big data and to scale these models, up to the Big Data, platforms.

Refer Slide Time :(0: 39)

What is Machine Learning?

- **Learning from data:**

- Machine learning is the field of study that focuses on computer systems that can learn from data. That is the system's often called models can learn to perform a specific task by analyzing lots of examples for a particular problem. For example, a machine learning model can learn to recognize an image of a cat by being shown lots and lots of images of cats.

What Characteristics Do Cats Have
TRIANGLE EARS AND NOSE
FUR
TAIL
WHISKERS
PAWS

Machine learning overview: What is a machine learning? Machine learning is the learning from the data. So, machine learning is the field, of study that focuses, on computer system that can learn from the data. That is the system's often called, the model can learn, to perform a specific task by analyzing, the lot of examples, for a particular problem. For example, machine learning model, can learn to recognize an

image, of a cat, by being shown a lots and lots of example and images, of the cat for example, cat can be characterized, by from the examples, that it has, the as a triangle ears and nose, it has a far, it has a tail, paws and whiskers. So, with these features, you can identify, characterize the cat.

Refer Slide Time :(1: 36)

What is Machine Learning ?

1. Learn from Data

2 **No explicit programming:** This notion of learning from data means that a machine learning model can learn a specific task without being explicitly programmed. In other words, the machine learning model is not given the step by step instructions on how to recognize the image of a cat.

• Instead, the model learns what features are important in determining whether it picture contains a cat from the data that has analyzed. Because the model learns to perform this task from data it's good to know that the amount and quality of data available for building the model are important factors in how well the model learns the task.

So, another feature of machine learning is. So, another aspect, which identifies or which distinguish machine learning is that it does not, have any programming. So, it has no explicit programming, involved in problem Solving. So, this much notion, of learning from the data means that, machine learning models, can learn a specific task, without being explicitly programmed. In other words, the machine learning models, is not given any step-by-step, instruction on how to recognize, the image of a, of a cat. Instead, the modern learns, with the features that are important, in determining whether the picture, contains the cat from the data, that has well analyzed, because the model learns to perform this task from the data, it is good to know, that the amount and quality of data available, for building the model are important factors in how, well the model learns, the task, therefore it does not have, any explicit programming and we have also, seen that, it machine learning is the learning from, from the data. So, there are two different aspects of a machine learning we have covered that is, it learns from the data and second thing is, it there is no explicit programming, involved in Solving the problem using machine learning.

Refer Slide Time :(3: 11)

What is Machine Learning ?

- **Discovering hidden patterns:** Because machine learning models can learn from data that can be used to discover hidden patterns and trends in the data.

- **Data-driven decisions:** These patterns and trends lead to valuable insights into the data. Thus the use of machine learning allows for data driven decisions to be made for a particular problem.

- So to summarize, the field of machine learning focuses on the study and construction of computer systems that can learn from data without being explicitly programmed. Machine learning algorithms and techniques are used to build models, to discover hidden patterns and trends in the data allowing for data-driven decisions to be made.
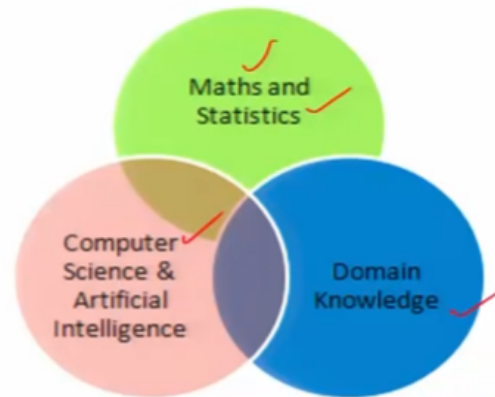
And third important characteristics of the machine learning is, about discovering the hidden patterns. Now, because machine learning models can learn from the data, that can be used to discover, hidden patterns and trends, in the data, which is hidden. So, the discovering hidden patterns is also, an important aspects of the big data, wipe of the machine learning, why because? It learns from the data. So, it can also discover the hidden patterns, which are embedded into the data and also, which and also captures, the trend, which is there inside the or within the data, important characteristics, which identifies or which characterizes, the machine learning is about the data driven decisions. So, these patterns and the trends, get to the valuable insight, into the data, thus the use of machine learning allows, the data driven applications or decisions to be made for a particular problem. Thus to summarize, the field of machine learning focuses on the study and the construction of computer system that can learn from the data, without being explicitly programmed, machine learning algorithm techniques are used to build models, to discover hidden pattern and to identify the trends, in the data, allowing for data, driven decisions to be made.

Refer Slide Time :(4: 47)

So, machine learning is an interdisciplinary field. So, in applying machine learning to a problem, the domain knowledge, is essential, to success of the end results, by domain knowledge we mean an understanding of the application or the business domain. In which we are solving the problem. So, the knowledge about the application the data related application and how, the outcomes will be used, are crucial in driving the process of building machine learning model. So, domain knowledge is also, an integral part of a machine learning solution, also we can see that, the machine learning is an interdisciplinary field, that is, it requires the understanding and the background of the roots of machine learning draws from applying the statistics and also, the computer science AI, these are computer science AI, statistics, I had also the domain knowledge, is the intersection of this or it's an interdisciplinary field of all three together, defines the machine learning.

Refer Slide Time :(6: 49)

Example Application of Machine Learning

- **Credit card fraud detection:** One application of machine learning that you likely used this past weekend, or even just today, is credit card fraud detection. Every time you use your credit card, the current purchase is analyzed against your history of credit card transactions to determine if the current purchase is a legitimate transaction or a potentially fraudulent one. If the purchase is very different from your past purchases, such as for a big ticket item in a category that you had never shown an interest in or when the point of sales location is from another country, then it will be flagged as a suspicious activity.
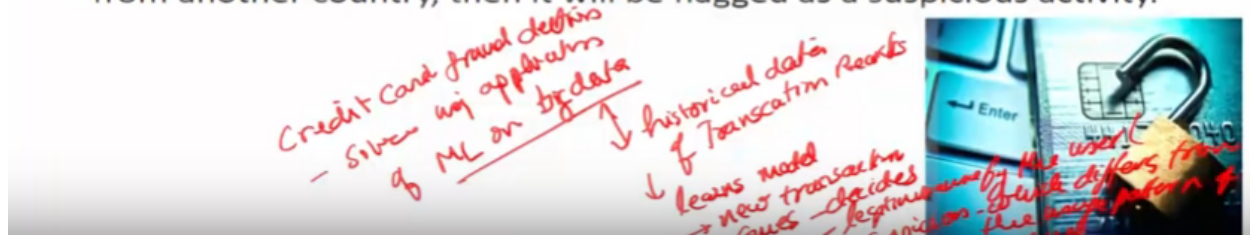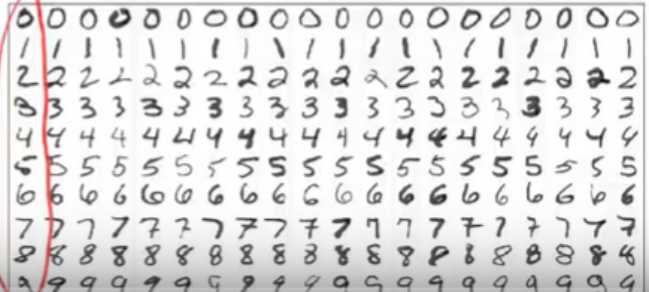
So, the example application of machine learning that means, where the machine learning can be applied, is in the credit card fraud detection, that is one of the machine learning, application is that you are likely to use, the in the weekend, or even today is the credit card, fraud detection. Every time you use the credit card, the current purchase is analyzed against your history of the car transaction to determine whether if the current purchase, is legitimate of the previous history transactions, or potentially a fraudulent one, I in this process the credit card, fraud can be detected by applying the machine learning, on this particular problem. So, if it is, did if it is, if the purchase is very, different from your past purchases, that learned from the machine learning, algorithm using, using the analysis of the data, such as for the big-ticket, item in the category that you had never, should an interest in, or when the point of sales is from another country then it will be flagged as the suspicious activity, therefore such applications like credit card, fraud detection, can be solved using, using an application, of machine learning on the big data. So, the big data is the historical data, of the transaction records. So, that means, from this particular historical data, of transaction records, the machine learning learns, the model and whenever a new, whenever a new transaction, comes then it decides, whether it's a legitimate, use by, the user or it is, suspicious that one's which differs, from the usage pattern of that user.

Refer Slide Time :(9: 57)

Now, another example is, about handwriting recognition. So, handwritten digit, recognition can be easily done with the help of machine learning. So, when you deposit a handwritten check, in an ATM, the machine learning process is used to read the numbers, written on the check, to determine the amount of the deposit handwritten digits are trickier, to be deciphered, then typed digit due to the due to many variation, in the people's handwriting. So, the so this handwritten. So, handwritten characters can be identified between zero to nine, the digits, using machine learning techniques. So, this handwritten, digit recognition is another such application of a machine learning.

Refer Slide Time :(10: 56)

Example Application of Machine Learning

- **Recommendations on websites:** After you buy an item on a website you will often get a list of related items. Often this will be displayed as customers who bought this item also bought these items, or you may also like.
- These related items have been associated with the item you purchased by a machine learning model, and are now being shown to you since you may also be interested in them.

Men's Black & Brown Canvas Sneakers
$112.00

RECOMMENDED FOR YOU

Men's Skinny Jeans

Men's Hoodie Sweatshirt

Men's Luxury Sunglasses

Men's Stylish Sweater

Now, further examples, of machine learning is, about recommendations on the website. So, after you, after you buy an item, on a website you will often get a list of related item. Often this will be displayed as the customers, who bought this item, also what these items are you may also, like that these related items have been associated with them, with the item you purchased by the machine learning model and now being shown you, shown to you since, you are also, interested to buy in them for example, when you purchase an item, let us say, as hoe then, you are also displayed there commended items, maybe based on your past purchases or the other customer who purchased this item, along with this item, what are the other items? They are purchased based on that, these recommendations are being made. So, the so, a particular item and, its associated items are being identified, by the machine learning and is to be recommended on the website, for online purchases, of the related items and these, recommendations are given to the customers. So, such kind of recommendations, on the website, is also an application of the machine learning.

Refer Slide Time :(12: 27)

More Applications of Machine Learning

- Targeted ads on mobile apps
- Sentiment analysis
- Climate monitoring
- Crime pattern detection
- Drug effectiveness analysis

Further applications of machine learning includes the targeted advertisements on the mobile apps, sentiment analysis, climate monitoring, crime pattern detection, drug effectiveness analysis.
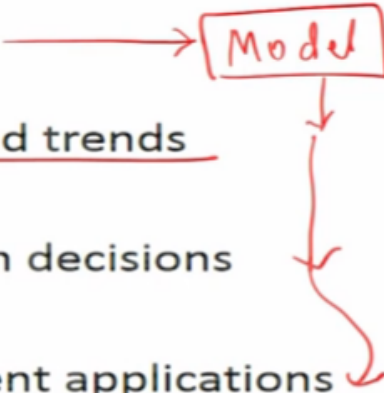
Refer Slide Time :(12: 41)

What's in a Name of Machine Learning?

- **Machine learning:** Machine learning has its roots since statistics, artificial intelligence, and computer science among other fields. Machine learning encompasses the algorithms and techniques used to learn from data.
- **Data mining:** The term data mining became popular around the time that the use databases became common place. So data mining was used to refer to activities related to finding patterns in databases and data warehouses. There are some practical data management aspects to data mining related to accessing data from databases. But the process of finding patterns in data is similar, and can use the same algorithms and techniques as machine learning.
- **Predictive analytics** refers to analyzing data in order to predict future outcomes. This term is usually used in the business context to describe activities such as sales forecasting or predicting the purchasing behavior of a customer.
- **Data science** is a new term that is used to describe processing and analyzing data to extract meaning. Again machine learning techniques can also be used here. Because the term data science became popular at the same time that big data began appearing, data science usually refers to extracting meaning from big data and so includes approaches for collecting, storing and managing big data.

Now, let us decipher, what's there in the name, behind the machine learning I and, other such terms, which are related terms, which are available. So, machine learning is, machine learning has its roots since, the statistics artificial intelligence and computer science, among other fields, together they have, contributed and the machine learning field. So, machine learning and compasses, algorithms are techniques, that is used to learn from the data. Now, a related field which is called a, 'Data Mining'. The data mining became popular, around the time, when the databases, were in common use, at databases and data warehouses, were at the main streams, in the business operations. So, at that time this, the data mining are of great importance and there are and there are, other practical data management aspect to the data mining related to accessing the data from the databases. But, the process of finding pattern in the data, is similar and similar to the machine learning and can be used and can and therefore, the same algorithms, which are there in data mining can also be used, as Some of the algorithmic techniques, in the machine learning that we will see. So, therefore data mining is, the field which was, around the databases and data warehouses and the therefore, there in that term, there in that technique, the algorithms and techniques are common, therefore some of these techniques from data mining and algorithms are also used in the machine learning. So, data mining, is differs from the machine learning in the sense that, it is, it was focusing around, the databases and data warehouses, data retrieval and finding the patterns within that, particular databases and data management. Now, the another term which is called a, 'Predictive Analytics. So, predictive analytics is the term which is used in the business houses. So, predictive analysis refers to analyzing the data, in order to predict the future outcomes this term, is usually used in the business context to determine, to scribe the activities such as, sales forecasting or predicting the purchase behavior of the customer. This particular predictive analytics, of analyzing the data also, uses the machine learning. But it is the, the it is the term, which is used to, predict the future outcomes and data mining is, to find the patterns in the databases and data warehouses and machine learning is, about the algorithms a technique used to learn from the data, furthermore the data science, is also related terms. So, data science is a new term, that is used to describe the processing and analyzing the data, to extract the meaning out of it, again the machine learning techniques can also be used here, in the data science, because the term data science, became popular, at the same time the big data, began appearing. So, data science usually refers to, extracting the meaning from the big data and so includes the approaches for collection, storing and managing the big data. So, therefore data science and also, uses the techniques of machine learning, to extract the meaning out of the data. Therefore all these, terms related to the, which are coin around the machine learning, we have just summarized, again we can say that, the machine learning is applied in data science and in predictive analytics, whereas the algorithms and techniques, which are used, in machine learning can be drawn from, the data mining. So, all are or they are all, related terms. So, we can say that, machine learning, can be applied, in data science, predictive analytics, whereas, the algorithms and techniques of, data mining is used in machine learning.

Refer Slide Time :(17: 49)

Machine Learning Models

- Learn from data
- Discover patterns and trends
- Allow for data-driven decisions
- Used in many different applications

So, let us see the machine learning, what do you mean by machine learning models? So, when it run learns from the data and discovers the patterns and trends, it allows also it allows the data-driven decisions, to be used in many applications. So, how to learn, from the data after learning from the data, it forms, the algorithm which is called a, 'Models'. So, we will see what do you mean by the model. So, it learns from the data, which we can also say that the model fits on that data. So, once the model is, identified then when a new data, is given it will discover, the patterns and the trends and this way, it allows, the data-driven decisions to be made and this particular way that means, learning from the data or this particular technique, is very useful, in many different applications, that we have seen, in the previous slides.

Refer Slide Time :(18: 58)

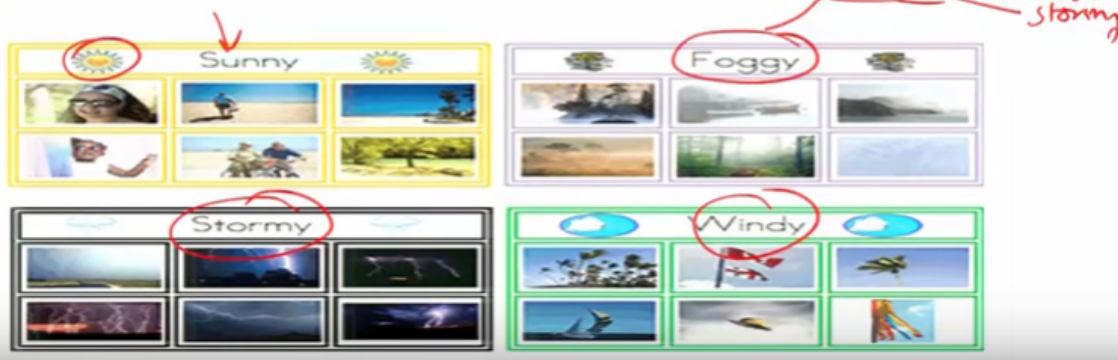**Categories Of Machine Learning Techniques**

- Classification
- Regression
- Cluster Analysis
- Association Analysis

Now, categories of machine learning techniques. So, machine learning techniques can be categorized, into four different types. So, first type is called the, 'Classification'. So, second type is called, 'Regression', 'Cluster Analysis' and 'Association Analysis'. Let us understand these categories, of machine learning techniques and then we will go, in more detail of, the different algorithms, which falls into these different categories.

Refer Slide Time :(19: 31)

Classification
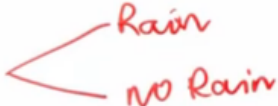
- **Goal:** Predict category
- In classification, the goal is to predict the category of the input data. An example of this is predicting the weather as being sunny, rainy, windy, or cloudy.

So, categories of machine learning, is the first one is called, 'Classification Technique' of 'Machine Learning'. So, here in this classification, the goal is to protect is the goal is to predict the category. So, in classification the goal is to predict the category, of category of the input data, for example, an example of this classification, is to predict the weather, as it is sunny, rainy, windy, or cloudy. So, the weather forecasts can be. So, weather condition can be a sunny or it can be a foggy, it can be a windy, or it can be stormy. So, this kind of categorization, that is the weather, which can be categorized, into sunny, windy, foggy and stormy. Is two that means given, this weather input, you can classify, into different categories, then it is called a, 'Classification Problem'.

Refer Slide Time :(20: 42)

Classification

- Some more examples of classification are classifying a tumor from a medical image as being benign or malignant.

- Predicting whether it will rain the next day.

- Determining if a loan application is high-risk, medium-risk or low-risk.

- Identifying the sentiment of a tweet or review as being positive, negative, or neutral.

 Some other examples, of classification a reclassifying a tumor, from the medical image as being, benign or malignant. So, predicting whether, it is, whether it will rain, the next day is also, a classification problem. So, that means, it is predicting whether it will, it will rain or it will, wilt rain or no rain. So, this becomes a binary classification problem. So, determining if the loan application is a high risk, medium risk and the low risk. So, if the loan application, can be classified, into three different categories, one of these three different categories, then it is also a classification problem. Now, identifying the sentiments, of a tweet or the review, as being positive, negative and neutral is also, is also, a classification problem. So, given a tweet and we can if, we can identify the sentiments, whether it is, the positive, negative or neutral, then this again, can be classified or this can also be, a classification problem of machine learning.

Refer Slide Time :(22:00)

 Now, let us see the regression, problem of machine learning. So, when your model, has to predict the numerical value, like in the, in the previous slide we have seen that, it is predicting the category. But, now if it is predicting a numerical value instead of category, then the task becomes, the regression problem. So, for example, an example of, regression is to predict the price, of a stock. So, the stock price, is a numerical value not a category. So, this is a regression task, instead of the classification task. So, if you are, if you were, to predict whether the stock price will rise or fall that means, it will classify as, rise or fall, then it would be a classification problem. But, if you are, finding out the price, of a or to predict the price of a stock then it becomes a regression problem. But, if you are predicting the actual price of that it is regression. So, that we have already, understood. So, that is the main difference between classification and regression, in classification, you are predicting at a category and in regression, you are predicting a value, a numeric value. So, in regression we are predicting a numeric value, whereas in classification problem, we are predicting category.

Refer Slide Time :(24: 06)

Regression Examples

- Estimating the demand of a product based on time or season of the year.

- Predicting a score on a test.

- Determining the likelihood of how effective a drug will be for a particular patient.

- Predicting the amount of rain for a region.

Now, further examples of, regression problems are estimating the demand of a product, based on the time or the season of a year. I am predicting a score of a test, determining the likelihood of how, effective a drug will be for a particular patient. I am predicting the amount of rain for a particular region.

Refer Slide Time :(24: 26)

Cluster Analysis of ML.

- In cluster analysis, the goal is to organize similar items in your data set into groups. A very common application of cluster analysis is referred to as customer segmentation. This means that you're separating your customer base into different groups or segments based on customer types.

- For example it would be very beneficial to segment your customers into seniors, adults and teenagers. These groups have different likes and dislikes and have different purchasing behaviors. By segmenting your customers to different groups you can more effectively provide marketing adds targeted for each groups particular interests. Note that cluster analysis is also referred to as clustering.

Business applications

Sales order
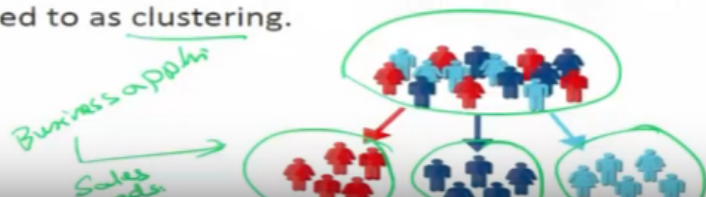
Now, the third category, of machine learning is called, 'Cluster Analysis'. So, let us see, about in cluster analysis the goal is to organize, similar items in your data, set into the groups. A very common application of cluster analysis, is referred to as the customer segmentation. This means that you are separating your customer base, into different groups and segments, based on customer types. For example, it would be very beneficial, to segment your customers into seniors, adults and teenagers. These groups have different likes and dislikes and I have different purchasing behaviors. By segmenting your customers to different groups you can, target more advertisements and to this particular group of, group having common interest. Don't that, the cluster analysis also referred as, the clustering. So, in this example, we have shown that, if it is a, group of people, we can segment them, or we can cluster them, into different groups, of having the common interest and these groups can further be used for various business, applications to target them for, the sales advertisements, having the common interest and so on.

Refer Slide Time :(25: 26)

Other examples, of the cluster analysis includes somehow are; identifying the areas of similar topography, such as desert region and grassy areas and mountain regions and categorizing, different type of tissues from the medical images, reminding different groups of weather pattern, such as snowy, dry and monsoon and also, discovering the hotspot for various type of, crime from the police records, therefore this classification, this grouping together, the common or similar data, items is called, 'Clustering'. And this cluster analysis is also called as a, 'Clustering'.

Refer Slide Time :(27: 10)

Now, another category of machine learning is called the, 'Association Analysis' of Machine Learning'. So, in association analysis the goal, is to come up with a set of rules, to capture association between, the items or the events. These rules are used to determine whether, the items are events occur together. So, a common application of Association analysis is known as Market Basket analysis, which is used to understand the customer purchasing behavior. So, for example, Association analysis can reveal the banking customers who have CDs, certificates deposits and also trending to be interested, in investment, in other, in other investment, instruments such as, money market accounts and so on. So, therefore the information, this association analysis, inside information can be used for cross selling. So, if you advertise money market, account to your customers and likely open, they will likely open, such an instrument that is already seen, in these applications.

Refer Slide Time :(28: 27)
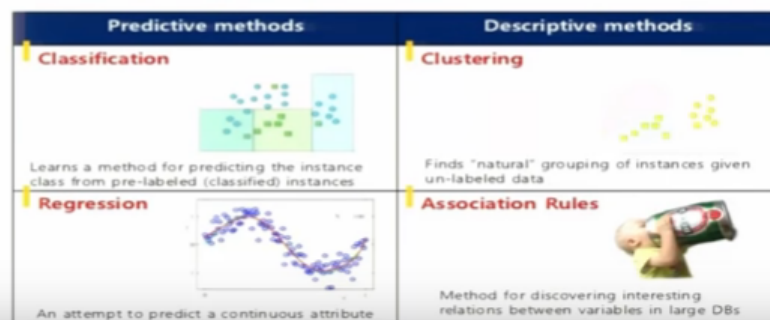
Association Analysis Examples

- Some other applications of association analysis are recommending similar items based on the purchasing behavior or browsing histories of customers.

- Finding items that are often purchased together, such as garden hose and potting soil, and offer sales on these related items at the same time to drive sales of both items.

- Identifying web pages that are often accessed together so that you can more efficiently offer up these related web pages at the same time.

So, association analysis other examples, are such as recommending items, which are similar items, based on the purchasing behavior or the browsing histories of customers. Finding items that are often purchased together and identifying the web pages that are of an, access together. So, that you can, more efficiently offer these web pages, at the same time.

Refer Slide Time :(28: 55)

Classification of Machine Learning Techniques

- So the different categories of machine learning techniques are classification, regression, cluster analysis, and association analysis.

Predictive methods	Descriptive methods
Classification	**Clustering**
Learns a method for predicting the instance class from pre-labeled (classified) instances	Finds "natural" grouping of instances given un-labeled data
Regression	**Association Rules**
An attempt to predict a continuous attribute	Method for discovering interesting relations between variables in large DBs

So, the classification of machine learning technique, we have different categories of machine learning techniques, scene let us say that, they are classification, regression, cluster analysis and Association analysis. And therefore, depending upon different problems, we have to apply, one of these different categories, to solve the problem, in the machine learning, landscape for example, for predictive methods, we can apply the classification or regression, depending upon whether the prediction of a category, it is there or a prediction of a value, similarly when the descriptive methods, are to be applied, then we have to choose, either between the clustering or between the Association rules.

Refer Slide Time :(29: 50)

Supervised vs. Unsupervised

- **In supervised approaches** the target, which is what the model is predicting, is provided. This is referred to as having labeled data because the target is labeled for every sample that you have in your data set.

- Referring back to our example of predicting a weather category of sunny, windy, rainy or cloudy, every sample in the data set is labeled as being one of these four categories. So the data is labeled and predicting the weather categories is a supervised task. In general, classification and regression are supervised approaches.

Now, furthermore the machine learning is also categorized as supervised and unsupervised. So, in supervised approaches the target, which is what the model is predicting is being provided. So, this is referred to as having the data, with a, with a label and because the target, is labeled, for every sample that you have, given in the dataset. So, referring back to our example, of predicting the weather category, of sunny, windy, rainy and cloudy, every sample of the weather data, is labeled as being one of these four categories. So, the data is labeled, by the domain expert and predicting the weather categories is also, called the, 'Supervised Task'. So, in general, the classification regression, both are the supervisor activity and approaches.

Refer Slide Time :(30: 48)

Supervised vs. Unsupervised

- **In unsupervised approaches** on the other hand, the target that the model is predicting is unknown or unavailable. This means that you have unlabeled data.

- Remember the cluster analysis example of segmenting customers into different groups. The samples in your data are not labeled with the correct group. Instead, the segmentation is performed using a clustering technique to group items based on characteristics that they have in common.

- Thus, the data is unlabeled and the task of grouping customers into different segments is an unsupervised one. In general, cluster analysis and association analysis are unsupervised approaches.

In unsupervised approach on the other hand, the target that the model is predicting is unknown or unavailable. This means that you have unlabeled data. Now, remember the cluster analysis example of, segmenting customers, into the different groups, of common of having similarities. So, the sample in your data, are not labeled, with the correct group, illustrate the segmentation, is performed using, the clustering techniques, to group items, based on the similarities or it is also called as a, 'Characteristics'. That they have, in common thus the data, is unlabeled, I on the task of grouping customers into different segments, is called, 'Unsupervised'. In general, the cluster analysis and Association analysis are unsupervised approaches.

Refer Slide Time :(30: 41)

Classification of Machine Learning Techniques
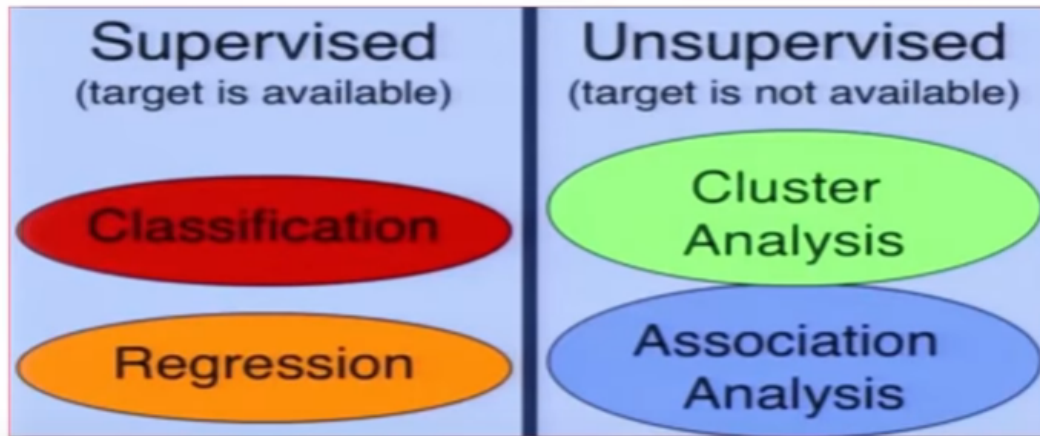
So, in short, we can classify all four techniques of a machine learning that is classification, regression as the supervised machine learning technique, whereas cluster analysis and Association analysis unsupervised, machine learning technique, depending upon, whether the target is available, with that data, set or not, if the target is available then, it is called, 'Supervised Learning'. If the target or label is not available, then it is called, 'Unsupervised Learning'. So, cluster analysis and Association analysis, there the target, is not available with the data set and hence, it is called, 'Unsupervised'. Various in classification and regression, the target or the label, is there with the data set, hence this activity of applying the dataset, is supervised machine learning technique.