

Machine Learning

ITM605

Unit 6

Introduction to Machine Learning

Application of Machine Learning

Steps in developing a Machine Learning Application.

Issues in Machine learning

Types of Machine Learning:

Supervised (Logistic Regression, Decision Tree, Support Vector Machine)

Unsupervised (K Means Clustering, Hierarchical Clustering, Association Rules)

Why Machine Learning?

- According to a recent study, machine learning algorithms are expected to **replace 25% of the jobs** across the world, in the next 10 years.
- With the rapid growth of big data and availability of programming tools like **Python and R** – machine learning is gaining mainstream presence for data scientists.
- Machine learning applications are highly **automated and self-modifying** which continue to improve over time with minimal human intervention as they learn with more data.

Machine Learning

- Learning is any process by which a system improves performance from experience.
- Machine Learning is concerned with computer programs that automatically improve their performance through experience.

What is Machine Learning?

- Discover new knowledge from large databases
- Ability to mimic human and replace tasks which require some intelligence
- That's why ML with big data is a deadly combination

Application of Machine Learning



Few Applications of Machine Learning

- Customer Analysis
- Fraud Detection

- Medical Screening and Diagnosis
- Face recognition
- Signature / fingerprint / iris verification
- Spam filtering
- Text translation
- Recommendation

Business Analytics Spectrum



Basic steps used in Machine Learning

There are 5 basic steps used to perform a machine learning task:

1. Collecting data:

2. Preparing the data:

3. Training a model: This step involves choosing the appropriate algorithm and representation of data in the form of the model. The cleaned data is split into two parts – train and test . Training data is used for developing the model. Test data is used for evaluation

4. Evaluating the model: To test the accuracy, test data is used. It is better to check accuracy of model on data which was not used during training.

5. Improving the performance: This step might involve choosing a different model altogether or introducing more variables

Issues in Machine learning

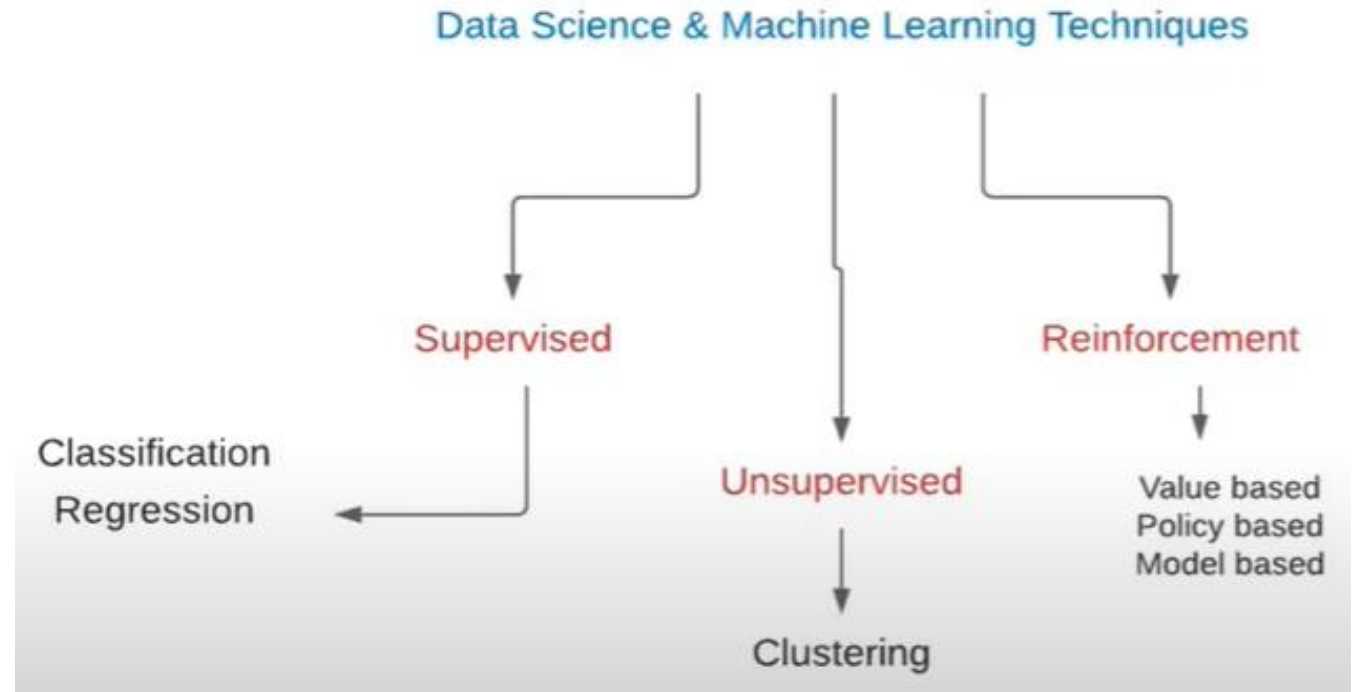
- Poor Quality of Data. ...
- Underfitting of Training Data. ...
- Overfitting of Training Data. ...
- Machine Learning is a Complex Process. ...
- Lack of Training Data. ...
- Slow Implementation. ...
- Imperfections in the Algorithm When Data Grows.

Issues in Machine learning

- What algorithms can approximate functions well (and when?)
- How does the number of training examples influence the accuracy?
- How does the complexity of the hypothesis representation impact it?
- How does noisy data influence the accuracy?
- What are the theoretical limits of learnability?
- How can prior knowledge of the learner help?
- What clues can we get from a biological learning system?
- How can systems alter their own representation?

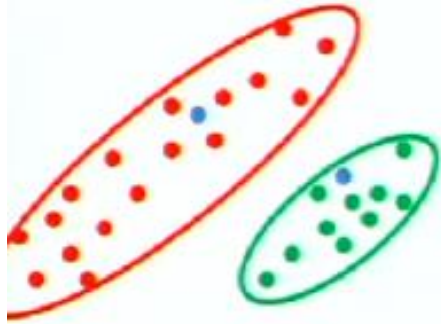
Types of Learning Model

- Supervised Learning Model
- Unsupervised Learning Model
- Reinforcement Learning Model

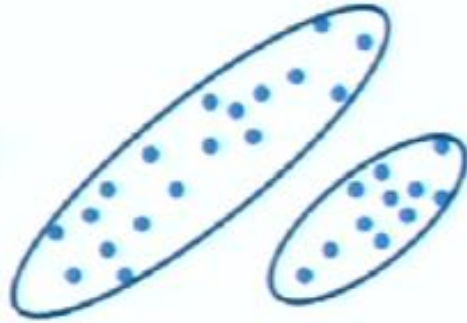


What Machine Learning consists of

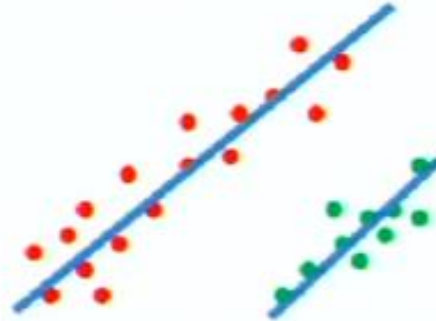
Classification



Clustering

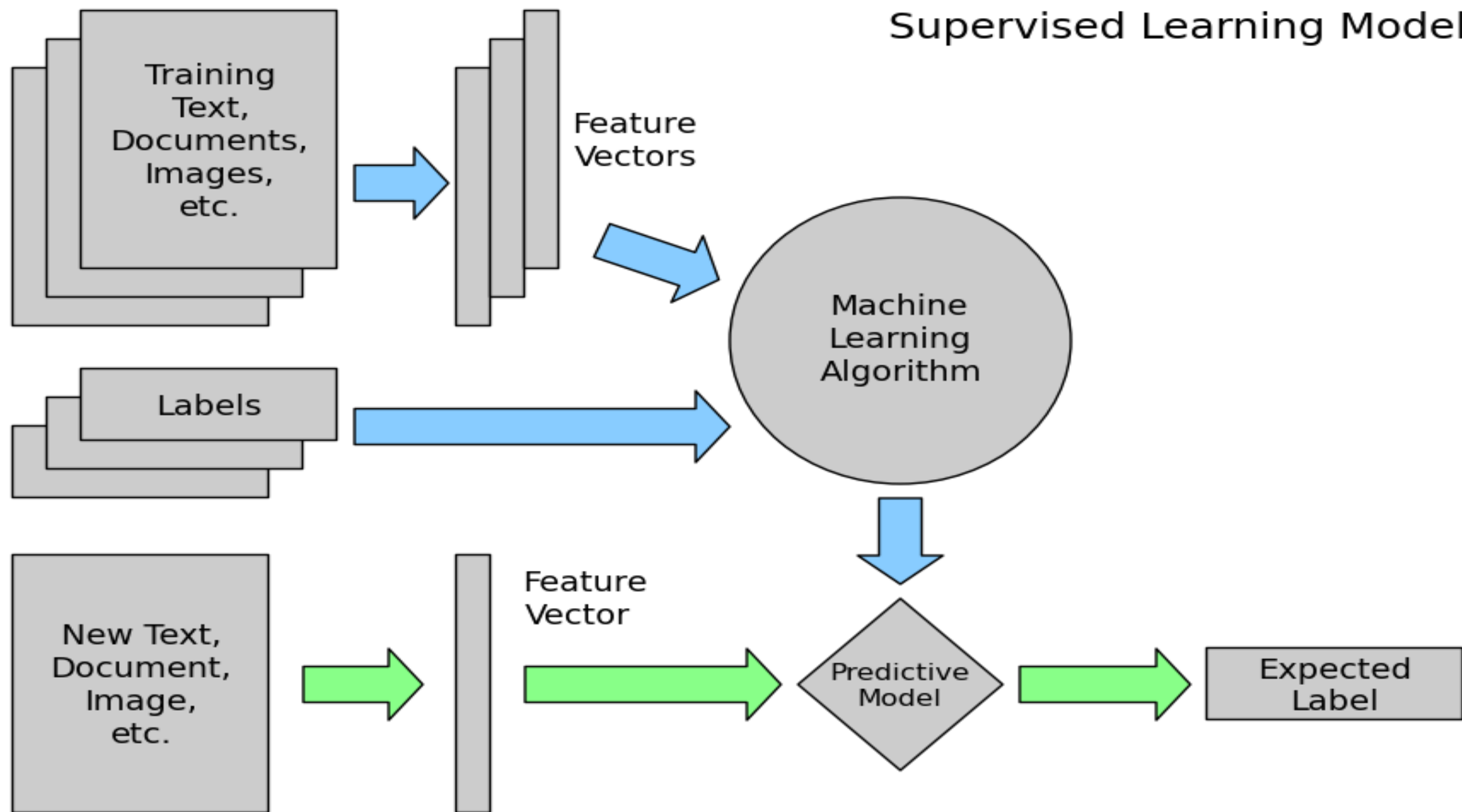


Regression

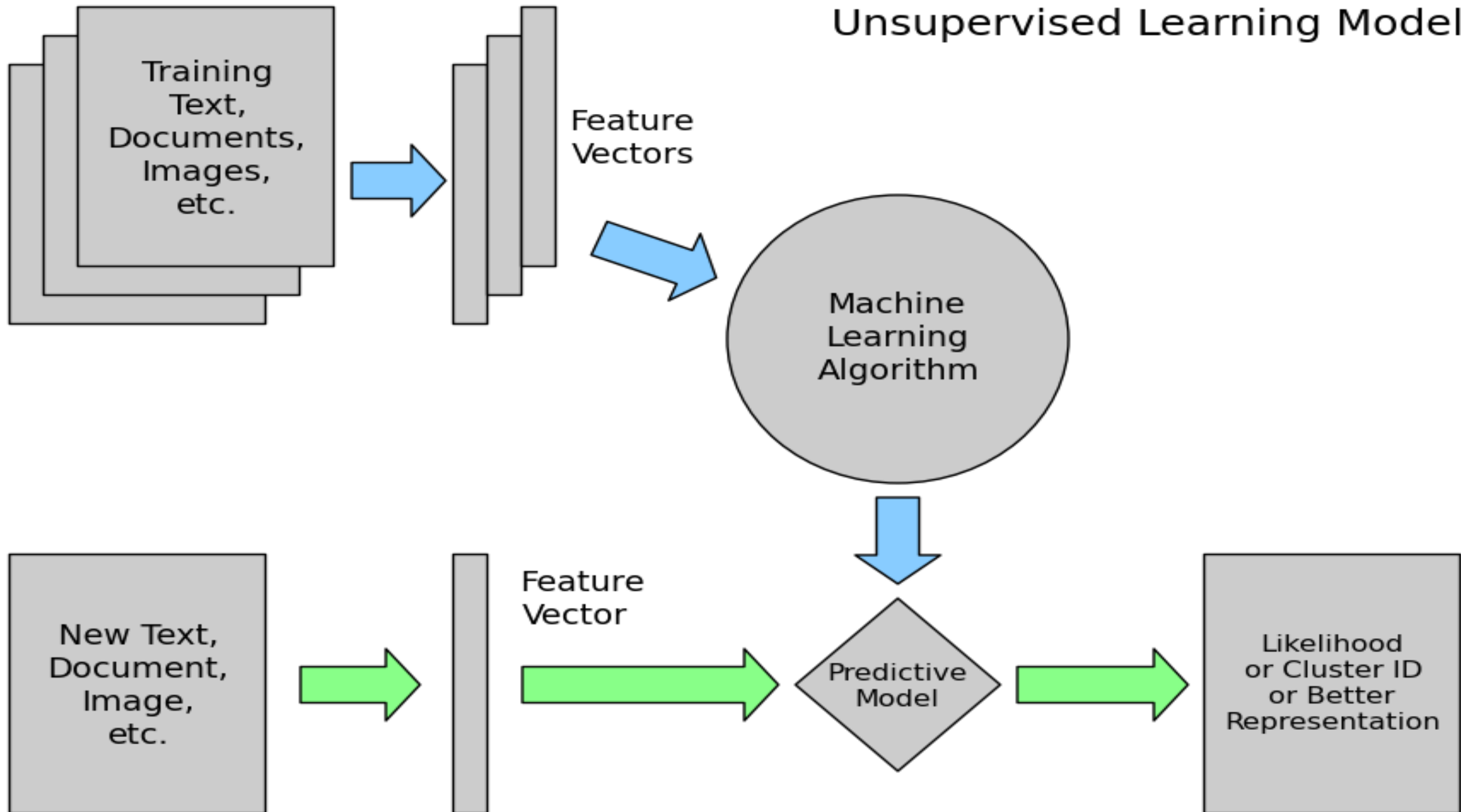


- Classification and clustering look similar
 - But the technology is quite different
- Classification and regression look quite different
 - But the technology is fairly similar

Supervised Learning Model



Unsupervised Learning Model



Few Applications of Machine Learning

- Customer Analysis
 - Fraud Detection
 - Medical Screening and Diagnosis
 - Face recognition
 - Signature / fingerprint / iris verification
 - Spam filtering
 - Text translation
 - Recommendation
-

Data Science & Machine Learning Relationship

- ☐ **Data Science:** complex study of the large amounts of data
 - ☐ Organizational Data
 - ☐ Structured or unstructured
 - ☐ Companies using Data Science: Amazon, Netflix, the healthcare sector, airlines etc.
- ☐ **Machine Learning:** capability to learn without being explicitly programmed
 - ☐ ML algorithms processes the data
 - ☐ Get trained for delivering future predictions, Ex. Facebook, Google, etc.

Classification Techniques

- Logistic regression
- Decision trees
- Naïve Bayes classifiers
- K-nearest neighbor
- Neural networks

Regression

Simple Linear Regression

Multiple Linear regression

Polynomial Regression

Clustering

K means Clustering

Hierarchical Clustering

Linear Regression- Objectives

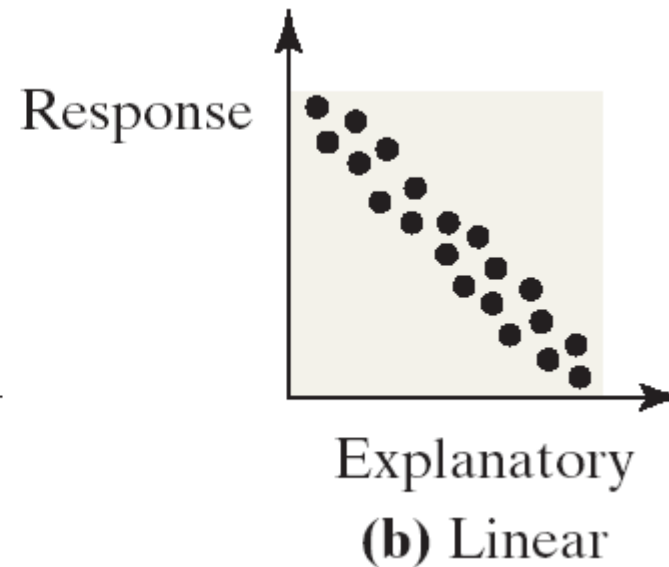
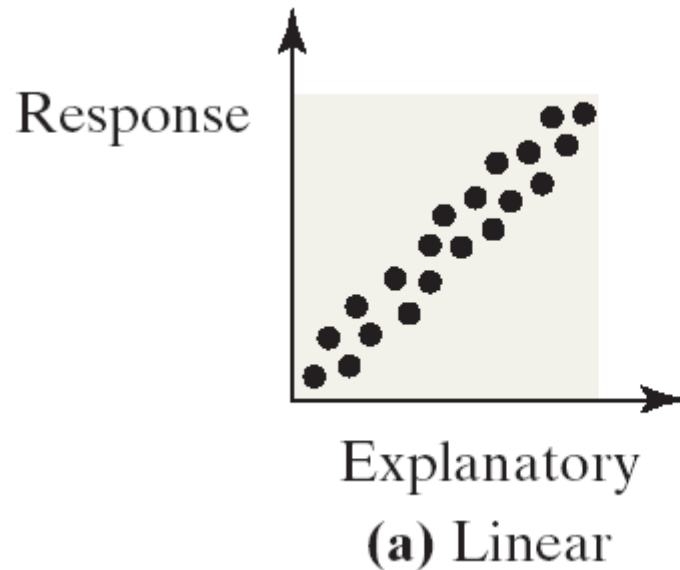
- To display the relationship between two quantitative variables with a best-fitted line called a **linear regression line**.

Types of Relation

- There are different types of relations between two variables
 - A relationship is linear when the points follow the general pattern of a line
 - A relationship is nonlinear when follow a general pattern, but it is not a line
 - A relationship has no correlation when the points do not show any pattern

Linear Correlations

- Linear relations or **linear correlations** have points that cluster around a line
- Linear relations can be either **positive** (the points slant upwards to the right) or **negative** (the points slant downwards to the right)



Techniques to discuss

Linear Regression

Analyzes relationship between response variable and the independent variable(s)
Simple Linear Regression:

$$y=a+bx$$

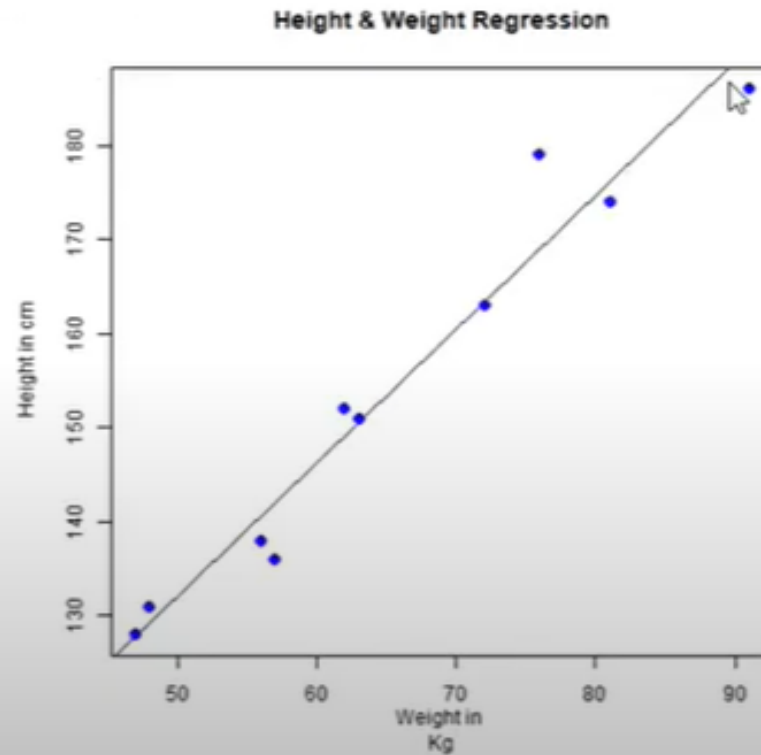
Some Examples:

- Relationship between height and weight
- Effect of fertilizer and water on crop yields
- Relationship between qualification and the salary of employee
- Relationship between height and age of individual
- Impact of product price on number of sales
- Impact of GPA on Placement of candidate

Techniques to discuss

Linear Regression

Linear regression assumes that there is linear relationship between dependent variable and independent variable.



Techniques to discuss

Logistic Regression:

- Linear regression is not capable of predicting probability
- Logistic Regression models the probability of some event
- Measures the relationship between the categorical dependent variable and one or more independent variables

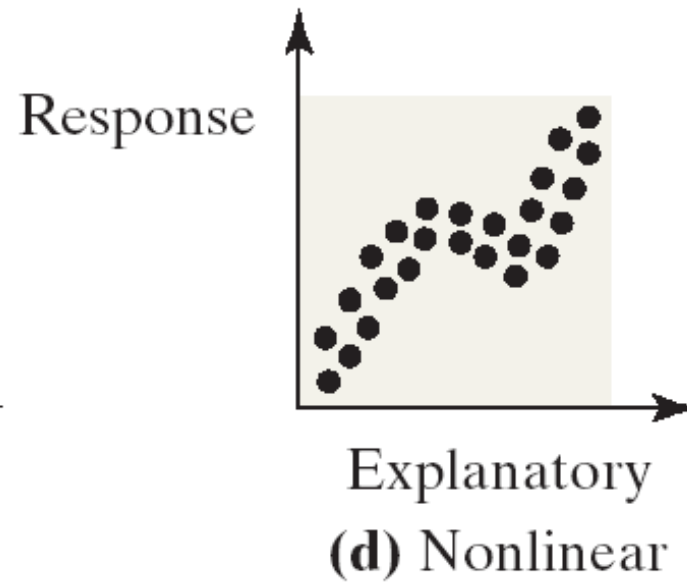
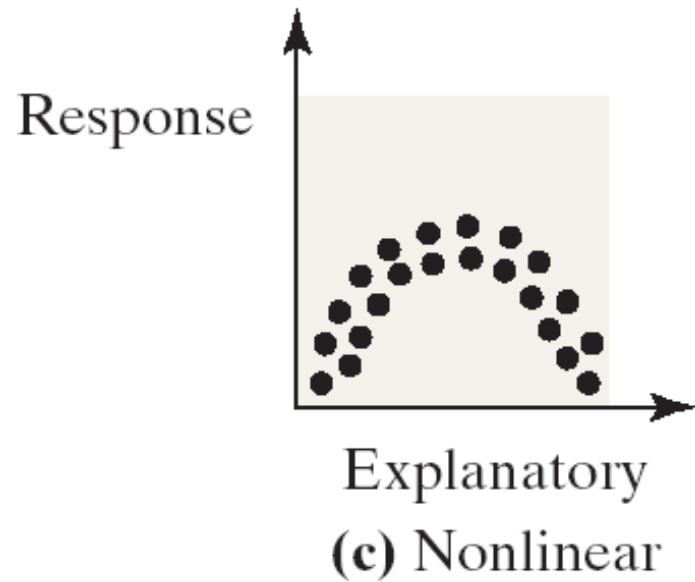
Ex. determining something is true or false, win or lose, spam or not spam, gender belongs to male or female etc.

- Used for classification
- Predict a qualitative response



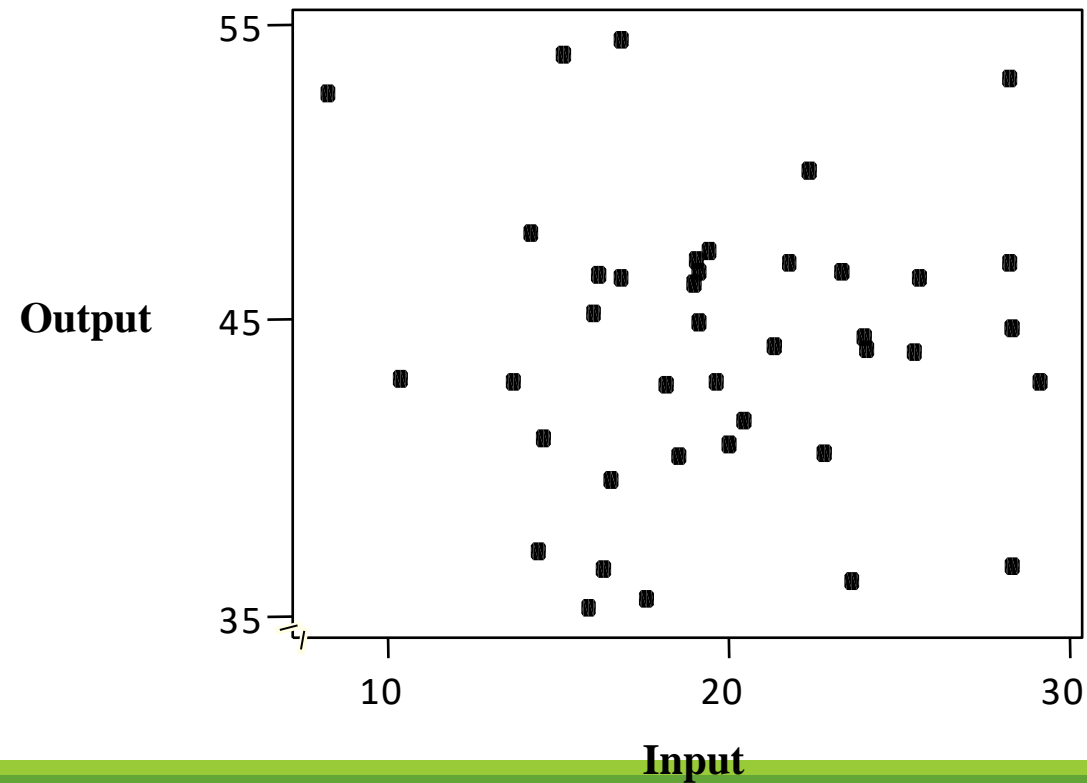
Nonlinear Correlations

Nonlinear relations have points that have a trend, but not around a line



No Correlation

As x increases, there is no definite shift in y :

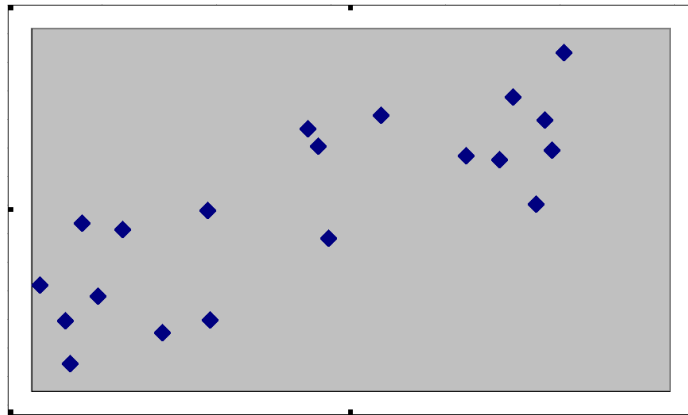


Linear Regression line prediction

- Find the regression line to fit the data and use the line to make predictions
- Compute the sum of squared residuals

Best Fitted Line

- If we have two variables X and Y which tend to be linearly correlated, we often would like to model the relation with a line that best fits to the data.



- We want to find the line that “best” describes the linear relationship ... the **regression line**

Multiple linear Regression

known as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable.

The goal is to model the linear relationship between the explanatory (independent) variables and response (dependent) variable.

Multi Regression

Formula and Calculation of Multiple Linear Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon$$

where, for $i = n$ observations:

y_i = dependent variable

x_i = explanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = the model's error term (also known as the residuals)

Tell You

Linear regression can only be used when one has two continuous variables—an independent variable and a dependent variable.

A multiple regression model extends to several explanatory variables.

MLR examines how multiple independent variables are related to one dependent variable.

Can I do a multiple regression by hand?

Probably not.

Multiple regression models are complex and become even more so when there are more variables included in the model or when the amount of data to analyse grows.

To run a multiple regression you will likely need to use specialized statistical software, or functions within business programs like Excel.

Consideration

In multiple linear regression, it is possible that some of the independent variables are actually correlated with one another

so it is important to check these before developing the regression model.

If two independent variables are too highly correlated ,then only one of them should be used in the regression model.

Use of multiple linear Regression

You can use multiple linear regression when you want to know:

The value of the dependent variable at a certain value of the independent variables (e.g. the expected yield of a crop at certain levels of rainfall, temperature, and fertilizer addition).

References

1. <https://www.educba.com/applications-of-machine-learning/>