
Postgraduate Certificate in Hydroinformatics in Civil Engineering

Data Analysis and Visualization

Abstract Data Type refers to a high-level concept in programming that allows for the definition of a data type in terms of its operations, without worrying about its implementation, in the context of data analysis and visualization, it is essential to understand abstract data types to effectively work with different data structures.

In data analysis, Abstract Syntax Tree is a tree representation of the source code, that can be used for analyzing and optimizing the code, it is essential to understand abstract syntax trees to work with programming languages and develop efficient algorithms.

Access Control is a security process that controls who can access or modify data in a system, in the context of data analysis and visualization, access control is crucial to prevent unauthorized access to sensitive information.

Accuracy is a measure of how close the results of a model are to the actual values, in the context of data analysis and visualization, accuracy is essential to evaluate the performance of a model, and it is often measured using metrics such as mean absolute error or mean squared error.

Actionable Insight refers to the information that can be used to make informed decisions, in the context of data analysis and visualization, actionable insights are the ultimate goal of analysis, as they provide a clear understanding of what actions to take to achieve a specific goal.

Activation Function is a mathematical function that is used to introduce non-linearity into a neural network, in the context of deep learning, activation functions are essential to enable the model to learn complex patterns in data.

Active Learning is a subfield of machine learning that involves actively selecting the most informative samples for labeling, in the context of machine learning, active learning is useful when the cost of labeling is high, and it can help to reduce the amount of labeled data required.

AdaBoost is a popular boosting algorithm that combines multiple weak models to create a strong model, in the context of machine learning, AdaBoost is useful for handling high-dimensional data and reducing overfitting.

Agent-Based Modeling is a modeling approach that involves simulating the behavior of autonomous agents, in the context of complex systems, agent-based modeling is useful for understanding the dynamics of complex systems and predicting their behavior.

Aggregation is the process of combining multiple values into a single value, in the context of data analysis, aggregation is essential to reduce the dimensionality of data and extract meaningful insights.

Algorithm is a set of instructions that is used to solve a specific problem, in the context of computer science, algorithms are essential to develop efficient solutions to complex problems, and they are widely used in data analysis and visualization.

Alternative Hypothesis is a hypothesis that is used as an alternative to the null hypothesis, in the context of statistical testing, alternative hypotheses are essential to evaluate the significance of a result, and they are often used to compare the performance of different models.

Analytic is a term that refers to the process of analyzing data to extract insights, in the context of data

analysis and visualization, analytics is essential to make informed decisions, and it involves using various techniques such as regression, clustering, and decision trees.

Anomaly Detection is the process of identifying data points that are significantly different from the rest of the data, in the context of data analysis, anomaly detection is essential to identify unusual patterns or outliers that may indicate errors or unusual behavior.

API is an Application Programming Interface that allows different systems to communicate with each other, in the context of software development, APIs are essential to enable the integration of different systems and services, and they are widely used in data analysis and visualization.

Area Under the Curve is a metric that is used to evaluate the performance of a model, in the context of machine learning, area under the curve is essential to compare the performance of different models, and it is often used to evaluate the accuracy of a model.

Array is a data structure that is used to store a collection of values, in the context of programming, arrays are essential to store and manipulate data, and they are widely used in data analysis and visualization.

Artificial Intelligence is a field of study that involves developing intelligent systems that can perform tasks that typically require human intelligence, in the context of computer science, artificial intelligence is essential to develop systems that can learn, reason, and interact with humans, and it is widely used in data analysis and visualization.

Artificial Neural Network is a type of machine learning model that is inspired by the structure and function of the human brain, in the context of deep learning, artificial neural networks are essential to learn complex patterns in data, and they are widely used in data analysis and visualization.

Asymptote is a line that a curve approaches as the input or output increases without bound, in the context of mathematics, asymptotes are essential to understand the behavior of functions, and they are often used to evaluate the performance of a model.

Asynchronous Processing is a type of processing that allows multiple tasks to be executed concurrently, in the context of programming, asynchronous processing is essential to improve the performance and efficiency of a system, and it is widely used in data analysis and visualization.

Autoencoder is a type of neural network that is used to learn a compact representation of data, in the context of deep learning, autoencoders are essential to reduce the dimensionality of data and extract meaningful insights.

Autoregression is a type of regression analysis that involves modeling the relationship between a variable and its past values, in the context of time series analysis, autoregression is essential to forecast future values, and it is often used to evaluate the performance of a model.

Average is a measure of the central tendency of a distribution, in the context of statistics, average is essential to understand the characteristics of a distribution, and it is often used to compare the performance of different models.

Backpropagation is an algorithm that is used to train artificial neural networks, in the context of deep learning, backpropagation is essential to update the weights and biases of a model, and it is widely used in data analysis and visualization.

Bagging is a technique that involves combining multiple models to improve the performance and robustness of a model, in the context of machine learning, bagging is essential to reduce overfitting, and it is often used to evaluate the performance of a model.

Bayes' Theorem is a mathematical formula that is used to update the probability of a hypothesis based on

new evidence, in the context of probability theory, Bayes' theorem is essential to evaluate the uncertainty of a model, and it is widely used in data analysis and visualization.

Bias-Variance Tradeoff is a fundamental concept in machine learning that involves balancing the tradeoff between bias and variance, in the context of machine learning, bias-variance tradeoff is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Big data refers to the large amounts of structured and unstructured data that are generated by various sources, in the context of data analysis and visualization, big data is essential to extract meaningful insights, and it is widely used in various industries.

Binary Classification is a type of classification problem that involves predicting one of two classes, in the context of machine learning, binary classification is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Binomial Distribution is a discrete probability distribution that models the number of successes in a fixed number of independent trials, in the context of statistics, binomial distribution is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Bootstrap Sampling is a technique that involves resampling with replacement to estimate the variability of a statistic, in the context of statistics, bootstrap sampling is essential to evaluate the uncertainty of a model, and it is widely used in data analysis and visualization.

Box Plot is a graphical representation that is used to display the distribution of a variable, in the context of statistics, box plots are essential to understand the characteristics of a distribution, and they are often used to compare the performance of different models.

Categorical data refers to data that can take on one of a limited number of distinct values, in the context of data analysis, categorical data is essential to extract meaningful insights, and it is widely used in various industries.

Centroid is the center of a cluster in a clustering algorithm, in the context of machine learning, centroids are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Chaos Theory is a field of study that involves understanding complex and dynamic systems, in the context of complex systems, chaos theory is essential to understand the behavior of complex systems, and it is widely used in data analysis and visualization.

Classification is a type of supervised learning problem that involves predicting a categorical label, in the context of machine learning, classification is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Clustering is a type of unsupervised learning problem that involves grouping similar objects together, in the context of machine learning, clustering is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

Coefficient of Determination is a metric that is used to evaluate the performance of a model, in the context of machine learning, coefficient of determination is essential to compare the performance of different models, and it is often used to evaluate the accuracy of a model.

Collinearity is a phenomenon that occurs when two or more variables are highly correlated, in the context of statistics, collinearity is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Complexity is a measure of the difficulty of a problem or a model, in the context of computer science,

complexity is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Composition is a technique that involves combining multiple models to create a new model, in the context of machine learning, composition is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Confidence Interval is a range of values that is likely to contain the true value of a parameter, in the context of statistics, confidence intervals are essential to evaluate the uncertainty of a model, and they are widely used in data analysis and visualization.

Confusion Matrix is a table that is used to evaluate the performance of a classification model, in the context of machine learning, confusion matrices are essential to compare the performance of different models, and they are often used to evaluate the accuracy of a model.

Conjugate Gradient is an optimization algorithm that is used to minimize a function, in the context of machine learning, conjugate gradient is essential to update the weights and biases of a model, and it is widely used in data analysis and visualization.

Constrained Optimization is a type of optimization problem that involves finding the minimum or maximum of a function subject to constraints, in the context of machine learning, constrained optimization is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Contingency Table is a table that is used to display the relationship between two categorical variables, in the context of statistics, contingency tables are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Convex Optimization is a type of optimization problem that involves finding the minimum or maximum of a convex function, in the context of machine learning, convex optimization is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Correlation is a measure of the relationship between two variables, in the context of statistics, correlation is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Cosine Similarity is a metric that is used to measure the similarity between two vectors, in the context of machine learning, cosine similarity is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Cross-Validation is a technique that involves splitting the data into training and testing sets to evaluate the performance of a model, in the context of machine learning, cross-validation is essential to evaluate the performance of a model, and it is widely used in data analysis and visualization.

Curve Fitting is a technique that involves finding a curve that best fits a set of data points, in the context of statistics, curve fitting is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Data Augmentation is a technique that involves generating new data points by applying transformations to the existing data, in the context of machine learning, data augmentation is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Data Mining is the process of automatically discovering patterns and relationships in large data sets, in the context of data analysis and visualization, data mining is essential to extract meaningful insights, and it is widely used in various industries.

Data Preprocessing is the process of preparing the data for analysis, in the context of data analysis and visualization, data preprocessing is essential to improve the quality and accuracy of the data, and it is widely used in various industries.

Data Visualization is the process of creating graphical representations of data to gain insights and understand the relationships between variables, in the context of data analysis and visualization, data visualization is essential to communicate the results of an analysis, and it is widely used in various industries.

Decision Boundary is the boundary that separates the classes in a classification problem, in the context of machine learning, decision boundaries are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Decision Tree is a type of machine learning model that is used for classification and regression problems, in the context of machine learning, decision trees are essential to extract meaningful insights, and they are widely used in data analysis and visualization.

Deep Learning is a type of machine learning that involves using artificial neural networks with multiple layers, in the context of machine learning, deep learning is essential to learn complex patterns in data, and it is widely used in data analysis and visualization.

Density Estimation is the process of estimating the underlying probability density function of a data set, in the context of statistics, density estimation is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Dependent Variable is the variable that is being predicted or explained in a regression analysis, in the context of statistics, dependent variables are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Determinant is a value that is used to describe the scaling effect of a matrix on a region of space, in the context of linear algebra, determinants are essential to understand the properties of matrices, and they are often used to evaluate the performance of a model.

Dimensionality Reduction is a technique that involves reducing the number of features or variables in a data set, in the context of machine learning, dimensionality reduction is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Discrete data refers to data that can take on only a limited number of distinct values, in the context of data analysis, discrete data is essential to extract meaningful insights, and it is widely used in various industries.

Discriminative Model is a type of machine learning model that is used for classification problems, in the context of machine learning, discriminative models are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Distance Metric is a function that is used to measure the distance between two points, in the context of machine learning, distance metrics are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Distributed Computing is a type of computing that involves using multiple computers to solve a problem, in the context of computer science, distributed computing is essential to improve the performance and efficiency of a system, and it is widely used in data analysis and visualization.

Eigenvalue is a scalar that is used to describe the amount of change of a linear transformation, in the context of linear algebra, eigenvalues are essential to understand the properties of matrices, and they are often used to evaluate the performance of a model.

Eigenvector is a vector that is used to describe the direction of a linear transformation, in the context of

linear algebra, eigenvectors are essential to understand the properties of matrices, and they are often used to evaluate the performance of a model.

EM Algorithm is an algorithm that is used to find the maximum likelihood estimate of a model, in the context of machine learning, EM algorithm is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Ensemble Learning is a type of machine learning that involves combining multiple models to improve the performance and robustness of a model, in the context of machine learning, ensemble learning is essential to evaluate the performance of a model, and it is widely used in data analysis and visualization.

Entropy is a measure of the uncertainty or randomness of a data set, in the context of information theory, entropy is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Epoch is a term that refers to a single pass through the entire data set, in the context of machine learning, epochs are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Error Analysis is the process of analyzing the errors that occur in a model, in the context of machine learning, error analysis is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Expectation-Maximization Algorithm is an algorithm that is used to find the maximum likelihood estimate of a model, in the context of machine learning, expectation-maximization algorithm is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Exponential Distribution is a continuous probability distribution that is often used to model the time between events, in the context of statistics, exponential distribution is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Exponential Smoothing is a technique that is used to forecast future values, in the context of time series analysis, exponential smoothing is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Feature Engineering is the process of selecting and transforming the features or variables in a data set, in the context of machine learning, feature engineering is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Feature Extraction is the process of extracting relevant features or variables from a data set, in the context of machine learning, feature extraction is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Feature Selection is the process of selecting the most relevant features or variables in a data set, in the context of machine learning, feature selection is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Feedforward Neural Network is a type of artificial neural network that is used for supervised learning problems, in the context of deep learning, feedforward neural networks are essential to learn complex patterns in data, and they are widely used in data analysis and visualization.

Filter is a technique that is used to remove noise or unwanted patterns from a data set, in the context of signal processing, filters are essential to improve the quality and accuracy of the data, and they are widely used in various industries.

Finite Mixture Model is a type of machine learning model that is used for clustering and density estimation,

in the context of machine learning, finite mixture models are essential to extract meaningful insights, and they are widely used in data analysis and visualization.

First-Order Logic is a type of logic that is used to reason about objects and their properties, in the context of artificial intelligence, first-order logic is essential to develop intelligent systems, and it is widely used in data analysis and visualization.

Fisher Information is a measure of the amount of information that a data set contains about a parameter, in the context of statistics, Fisher information is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Fitting is the process of adjusting the parameters of a model to fit the data, in the context of machine learning, fitting is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Fixed Effects Model is a type of statistical model that is used to analyze the relationship between a dependent variable and one or more independent variables, in the context of statistics, fixed effects models are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Frequency Domain is a domain that represents the frequency components of a signal, in the context of signal processing, frequency domain is essential to analyze and understand the properties of a signal, and it is widely used in various industries.

Functional Programming is a programming paradigm that emphasizes the use of pure functions and immutable data structures, in the context of computer science, functional programming is essential to develop efficient and scalable systems, and it is widely used in data analysis and visualization.

Gaussian Distribution is a continuous probability distribution that is often used to model the distribution of a variable, in the context of statistics, Gaussian distribution is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Gaussian Mixture Model is a type of machine learning model that is used for clustering and density estimation, in the context of machine learning, Gaussian mixture models are essential to extract meaningful insights, and they are widely used in data analysis and visualization.

Generalization is the ability of a model to perform well on new, unseen data, in the context of machine learning, generalization is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Generative Model is a type of machine learning model that is used to generate new data points, in the context of machine learning, generative models are essential to extract meaningful insights, and they are widely used in data analysis and visualization.

Geospatial Analysis is the process of analyzing and understanding the relationships between data that has spatial components, in the context of data analysis and visualization, geospatial analysis is essential to extract meaningful insights, and it is widely used in various industries.

Gradient Descent is an optimization algorithm that is used to minimize a function, in the context of machine learning, gradient descent is essential to update the weights and biases of a model, and it is widely used in data analysis and visualization.

Gradient Boosting is a type of ensemble learning algorithm that is used to combine multiple models, in the context of machine learning, gradient boosting is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Graph is a non-linear data structure that consists of nodes and edges, in the context of computer science, graphs are essential to represent complex relationships between objects, and they are widely used in data analysis and visualization.

Graph-Based Algorithm is an algorithm that is used to analyze and understand the structure of a graph, in the context of computer science, graph-based algorithms are essential to extract meaningful insights, and they are widely used in data analysis and visualization.

Histogram is a graphical representation that is used to display the distribution of a variable, in the context of statistics, histograms are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Hyperparameter is a parameter that is used to control the behavior of a model, in the context of machine learning, hyperparameters are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Hypothesis Testing is a statistical technique that is used to test a hypothesis about a population, in the context of statistics, hypothesis testing is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Independent Component Analysis is a technique that is used to separate a multivariate signal into its independent components, in the context of signal processing, independent component analysis is essential to analyze and understand the properties of a signal, and it is widely used in various industries.

Independent Variable is the variable that is used to predict or explain a dependent variable, in the context of statistics, independent variables are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Inference is the process of drawing conclusions about a population based on a sample of data, in the context of statistics, inference is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Information Gain is a measure of the amount of information that a feature or variable provides about a target variable, in the context of machine learning, information gain is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Instance-Based Learning is a type of machine learning that involves storing and retrieving instances of data to make predictions, in the context of machine learning, instance-based learning is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

Instrumental Variable is a variable that is used to identify the causal effect of a treatment on an outcome, in the context of statistics, instrumental variables are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Interaction Term is a term that is used to model the interaction between two or more variables, in the context of statistics, interaction terms are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Interpolation is the process of estimating the value of a function at a point that is not in the data, in the context of mathematics, interpolation is essential to analyze and understand the properties of a function, and it is widely used in various industries.

Interquartile Range is a measure of the spread of a distribution, in the context of statistics, interquartile range is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

Intrinsic Dimensionality is the number of independent variables that are required to describe a data set, in the context of machine learning, intrinsic dimensionality is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Inverse Problem is a problem that involves estimating the parameters of a model based on observed data, in the context of machine learning, inverse problems are essential to extract meaningful insights, and they are widely used in data analysis and visualization.

Isomap is a technique that is used to reduce the dimensionality of a data set, in the context of machine learning, Isomap is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Iteration is the process of repeating a set of instructions or operations, in the context of computer science, iteration is essential to develop efficient and scalable systems, and it is widely used in data analysis and visualization.

Jacobian Matrix is a matrix that is used to describe the derivative of a function, in the context of mathematics, Jacobian matrices are essential to analyze and understand the properties of a function, and they are widely used in various industries.

Kalman Filter is an algorithm that is used to estimate the state of a system based on noisy data, in the context of signal processing, Kalman filter is essential to analyze and understand the properties of a signal, and it is widely used in various industries.

Kernel is a function that is used to map a data set into a higher-dimensional space, in the context of machine learning, kernels are essential to improve the performance and robustness of a model, and they are widely used in data analysis and visualization.

Kernel Density Estimation is a technique that is used to estimate the underlying probability density function of a data set, in the context of statistics, kernel density estimation is essential to understand the characteristics of a distribution, and it is often used to evaluate the performance of a model.

K-Means Clustering is a type of unsupervised learning algorithm that is used to group similar objects together, in the context of machine learning, K-Means clustering is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

K-Nearest Neighbors is a type of supervised learning algorithm that is used to classify or regress a target variable, in the context of machine learning, K-Nearest Neighbors is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

L1 Regularization is a type of regularization technique that is used to reduce the magnitude of the coefficients of a model, in the context of machine learning, L1 regularization is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

L2 Regularization is a type of regularization technique that is used to reduce the magnitude of the coefficients of a model, in the context of machine learning, L2 regularization is essential to improve the performance and robustness of a model, and it is widely used in data analysis and visualization.

Lagrange Multiplier is a technique that is used to optimize a function subject to constraints, in the context of machine learning, Lagrange multipliers are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Latent Variable is a variable that is not directly observed but is inferred from other variables, in the context of statistics, latent variables are essential to understand the characteristics of a distribution, and they are often used to evaluate the performance of a model.

Least Squares is a technique that is used to estimate the parameters of a linear model, in the context of statistics, least squares is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Leave-One-Out Cross-Validation is a technique that involves leaving out one example at a time to evaluate the performance of a model, in the context of machine learning, leave-one-out cross-validation is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Linear Algebra is a branch of mathematics that deals with the study of linear equations and vector spaces, in the context of mathematics, linear algebra is essential to analyze and understand the properties of linear systems, and it is widely used in various industries.

Linear Discriminant Analysis is a type of supervised learning algorithm that is used to classify a target variable, in the context of machine learning, linear discriminant analysis is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

Linear Regression is a type of supervised learning algorithm that is used to predict a continuous target variable, in the context of machine learning, linear regression is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

Local Outlier Factor is a measure of the degree to which an object is an outlier, in the context of machine learning, local outlier factor is essential to identify unusual patterns or outliers that may indicate errors or unusual behavior.

Logistic Regression is a type of supervised learning algorithm that is used to classify a binary target variable, in the context of machine learning, logistic regression is essential to extract meaningful insights, and it is widely used in data analysis and visualization.

LogLikelihood is a measure of the goodness of fit of a model, in the context of statistics, log-likelihood is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Loss Function is a function that is used to measure the difference between the predicted and actual values of a target variable, in the context of machine learning, loss functions are essential to evaluate the performance of a model, and they are often used to compare the performance of different models.

Machine Learning is a field of study that involves developing algorithms and statistical models that enable machines to perform tasks without being explicitly programmed, in the context of computer science, machine learning is essential to develop intelligent systems, and it is widely used in data analysis and visualization.

Mahalanobis Distance is a measure of the distance between a point and the center of a multivariate distribution, in the context of statistics, Mahalanobis distance is essential to evaluate the performance of a model, and it is often used to compare the performance of different models.

Manifold Learning is a type of unsupervised learning algorithm that is used to reduce the dimensionality of a data set, in the context of machine learning, manifold learning is