



INSAID INSIGHTS



INSAID

A Z

GLOSSARY OF DATA SCIENCE & MACHINE LEARNING

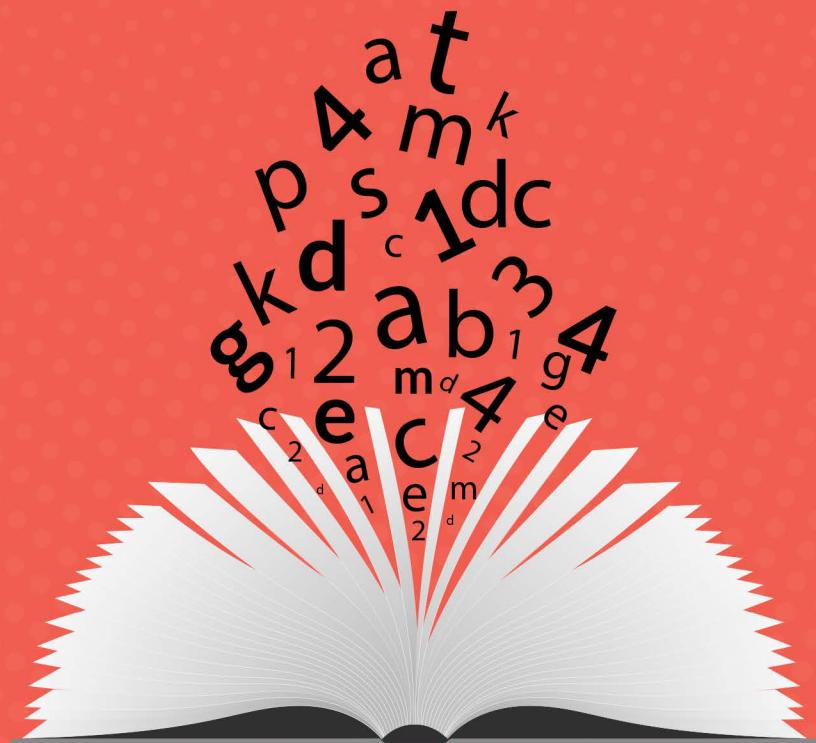


Table of Contents

A

Accuracy
Activation function
AdaGrad
AUC (Area Under the ROC Curve)
Algorithm
Artificial Intelligence

B

Backpropagation or 'backprop.'
Baseline
Batch
Bayes' theorem
Bayesian Network or Bayes Net
Bias term or bias
Binary classification
Big Data
Bucketing

C

Calibration layer
Candidate Sampling
Checkpoint
Chi-square test
Classification or class
Class-imbalanced dataset
Classification threshold
Clustering
Coefficient
Computational linguistics or Natural language processing or NLP
Confidence interval
Confusion matrix

D

Continuous variable
Convergence
Convex function
Correlation
Covariance
Cross-entropy

Data-driven Documents or D3
Data mining
Dataset
Data science
Data structure
Data wrangling
Decision boundary
Decision trees
Deep learning or deep model
Dense feature
Dependent variable
Dimension reduction
Discrete feature or discrete variable
Dropout regularization
Dynamic model

E

Early stopping
Embeddings
Empirical Risk Minimization (ERM)
Ensemble
Estimator
Example

F

False Negative or FN
False Positive or FP
False positive rate
Feature
Feature columns
Feature cross
Feature engineering
Feature set

G

GATE or General Architecture for Text Engineering
Generalization
Generalized linear model
Gradient
Gradient boosting
Gradient clipping
Gradient descent
Graph

H

Heuristic
Hidden Layer
Hinge loss
Histogram
Holdout data
Hyperparameter

I

Independently and identically distributed (IID)
Inference
Input layer
Inter-rater agreement

K

Kernel Support Vector Machines (KSVMs)
K-means clustering
K-nearest neighbors or kNN

L

Latent variable
Label
Layer
Lift
Linear regression
Logistic regression

M

Machine learning
Markov chain
Matrix
Mean
Mean Absolute error
Mean Squared error of MSE
Median
Mode
Model
Monte Carlo method
Moving average
Multivariate analysis

N

N-gram
Naive Bayes classifier
Natural language processing
Neural network
Normal distribution
Null hypothesis

O

Objective function
One hot encoding
Ordinal variable
Outlier
Overfitting

P

P value
Perceptron
Pivot table
Poisson distribution
Predictive analytics
Precision and recall
Predictor variables
Principal component analysis
Prior distribution

Q

Quantiles and quartiles

R

R
Random forest
Range
Regression
Reinforcement learning
Response variable
Ridge regression
Root Mean Squared Error or RMSE

S

S curve
Scalar
Semi-supervised learning
Serial correlation

Skewness
Spatiotemporal data
Standard deviation
Standard error
Standard normal distribution
Standardized score
Strata
Supervised learning
Support vector machine or SVM

T

T-distribution
Type I error
Type II error
T-test

U

Univariate analysis
Unsupervised learning

V

Variance
Vector
Vector-space

W

Weka

About INSAID

INSAID is India's first academic institution dedicated to research and education in Artificial Intelligence & Data Science. INSAID's mission is to groom professionals into Data leaders of tomorrow.



[DOWNLOAD NOW](#)



Welcome to another exciting book from INSAID Research Team!

From cancer research to online product recommendations, Data Science & Machine Learning today are not only driving billions of dollars in economic value but also solving problems, the solutions to which we didn't have till now.

Millions of jobs are being created thanks to power of Machine Learning. This excitement is driving companies across industries to recruit top talent and pay top dollars to professionals with skills in these areas.

And the whole reason you are probably reading this book is because you're excited about learning more about Data Science & Machine Learning.

Our goal through the INSAID INSIGHTS series is to help you take first steps in Data Science & Machine Learning.

Each book in the INSAID INSIGHTS series covers one specific area and we recommend that you check out other books in this series.

Also, in case you haven't already, do check out our blog. We have some amazing [articles, tutorials](#) that you wouldn't want to miss.

We would love to hear from you what new insights you gained from this book. Send us your feedback and love about this book to info@insaid.co.

All the best.

- INSAID Data Science Research Team

A

• • •

ACCURACY

Accuracy is the fractional number of correct predictions made by a 'classification model.' Accuracy = (No. of correct predictions/Total no. of examples), as per 'multi-class classification.'

ACTIVATION FUNCTION

It is a function that generates and passes an output value (usually nonlinear) to the next layer, by taking in the weighted sum of all the inputs from the previous layer. 'ReLU' or 'Sigmoid' are the examples of activation functions.

ADAGRAD

AdaGrad is an advanced gradient descent algorithm that rescales the gradients of each parameter. It allows each parameter to have an independent learning rate.

AUC (AREA UNDER THE ROC CURVE)

The area under the ROC curve represents that a classifier will be more confident that 'a randomly chosen positive example is actually positive' than 'a randomly chosen negative example is positive.'

AUC is an evaluation metric that considers all the possible classification thresholds.

ALGORITHM

An algorithm is a series of repeatable steps for executing a specific type of task with the given data. It is a process governed by a set of rules, to be followed by a computer to perform operations on the data.

A, B

...

ARTIFICIAL INTELLIGENCE

Artificial intelligence or AI is the ‘machines’ acting with apparent intelligence. Modern AI employs statistical and predictive analysis of large amounts of data to ‘train’ the computer systems to make decisions, that appear as intelligence.

BACKPROPAGATION OR ‘BACKPROP.’

Backpropagation is the primary algorithm for implementing ‘gradient descent’ on ‘neural networks.’ In a backprop algorithm, the output values of each node are calculated in a forward pass. Then, the partial derivative of the error corresponding to each parameter is calculated in a backward pass through the graph.

BASELINE

A baseline is a reference point for comparing ‘how well a given model is performing.’ It is a simple model that help model developers quantify the ‘minimal’ expected performance by a system for a particular problem.

BATCH

A batch is a set of examples used in ‘one gradient update’ or iteration of ‘model training.’

Similarly, a ‘batch size’ represents the number of examples in a batch.

B

...

BAYES' THEOREM

Bayes' theorem describes the probability of an event, based on prior knowledge of conditions that might be related to the event. For an observed outcome, Bayes' theorem describes that the conditional probability of each of the set of possible causes can be computed from the knowledge of the probability of each cause and the underlying conditional probability of the outcome of each event.

Mathematically, Bayes' theorem says, $P(A|B) = (P(B|A)P(A))/P(B)$

BAYESIAN NETWORK OR BAYES NET

A Bayesian network is used for reasoning or decision making in the face of uncertainty. It consists of graphs that represent the relationship between random variables for a particular problem. The reasoning in Bayes net depends heavily upon the Bayes' rule.

BIAS TERM OR BIAS

A bias term is an intercept or the offset from an 'origin.' Bias is represented as 'b' or 'w0.'

BINARY CLASSIFICATION

A binary classification is a type of classification task that outputs one of the two mutually exclusive conditions as a result. For example, a machine learning model either outputs 'Spam' or 'Not spam' if it evaluates the email messages.

B

...

BIG DATA

Big data corresponds to extremely large datasets, that had been 'impractical' to use before because of their 'volume,' 'velocity' and 'variety.' Such datasets are analyzed computationally to reveal patterns, trends, associations & conditional relations, especially relating to human behavior and interactions.

Crunching down such extensive data requires data science skills to reveal useful insights and patterns hidden from general human intelligence.

BUCKETING

Bucketing is the conversion of continuous features, based on their value range, into multiple binary features called 'buckets' or 'bins.' Instead of using a variable as a continuous 'floating-point' feature, its value ranges can be chopped down to fit into discrete buckets.

For example, given a temperature data, all temperatures ranging from 0.0 to 15.0 can be put into one bin or bucket, 15.1 to 30.0 into another and so on.

C

...

CALIBRATION LAYER

A calibration layer is used in post-prediction adjustment to minimize the 'Prediction bias.' The calibrated predictions and probabilities must match the distribution value of an observed set of labels.

CANDIDATE SAMPLING

Candidate sampling is the 'training-time' optimization method, in which the probability is calculated for all the positive labels, but only for 'random' samples of negative labels. The idea is based on an empirical observation that 'negative classes' can learn from 'less frequent negative reinforcement' as long as the 'positive classes' always get 'proper positive reinforcement.'

For example, if we use an example labeled as 'Ferrari', and the 'car' candidate sampling computes the predicted probabilities and the corresponding loss terms for the 'Ferrari' and 'Car' class outputs, in addition to random subset of other remaining classes such as 'Trucks', 'Aircraft' and 'Motorcycles.'

CHECKPOINT

The 'captured state of variables' of a model at a given time is called 'Checkpoint' data. Checkpoint data enables performing training across multiple sessions. It also aids in 'exporting' model weights. Checkpoint data also enables to continue task preemption training.

CHI-SQUARE TEST

Chi-square test is an analytical method to determine 'whether the classification of data can be attributed to some underlying law or chance.' The chi-square analysis is used to estimate whether two variables in a 'cross-tabulation' are correlated. It is a test to check for the 'independence' or 'degree of freedom' of variables.

C

• • •

CLASSIFICATION OR CLASS

Classification is used to determine the categories to which an item belongs. It is an example of classic machine learning task. The two types of classifications are 'binary classification' and 'multiclass classification.'

Example of a binary classification model is where a system detects 'spam' or 'not spam' emails.

Example of a multiclass classification is where a model identifies 'cars,' the classes being 'Ferrari,' 'Porsche,' 'Mercedes' and so on.

CLASS-IMBALANCED DATASET

The class-imbalanced dataset is a binary classification problem in which two classes have wide frequency gap.

For example, a viral flu dataset in which 0.0004 of examples have positive labels and 0.9996 examples have negative labels pose a class-imbalance problem.

Whereas, in 'a marriage success predictor' in which 0.55 of examples label the 'couple' keeping a long-term marriage and 0.45 examples label the 'couple' ending up in divorce, is 'not' a classification problem.

CLASSIFICATION THRESHOLD

A classification threshold is a scalar-value that is used when mapping 'logistic regression' results to binary classification. This threshold value is applied to a model's predicted score to separate the 'positive class' from 'negative class.'

For example, consider a logistic regression model, with a classification threshold value of 0.8, which estimates the probability of a given email message as being spam or not spam. The logistic regression values above 0.8 will be classified as 'spam' and values below 0.8 are classed as 'not spam.'

C

• • •

CLUSTERING

Clustering is an unsupervised algorithm for dividing ‘data instances’ into groups and not a ‘predetermined set of groups.’ the groups ‘clustered’ by the execution of such an algorithm are ‘because of similarities found amongst the instances.’

‘Centroid’ is the term used to denote the center of each such cluster.

COEFFICIENT

A coefficient is the ‘multiplier value’ prefixed to a variable. It can be a number or an algebraic symbol. Data statistics involve the usage of specific coefficient terms such as Cramer’s coefficient and Gini coefficient.

COMPUTATIONAL LINGUISTICS OR NATURAL LANGUAGE PROCESSING OR NLP

Computational linguistics or NLP is a branch of computer science to analyze the text of spoken languages like Spanish or English, and convert it into structured data that can be used to drive the program logic.

For example, a model can analyze and process text documents, Facebook posts, etc. to mine for potentially valuable information.

CONFIDENCE INTERVAL

The confidence interval is a specific range around a ‘prediction’ or ‘estimate’ to indicate the scope of error by the model. The confidence interval is also combined with the probability that a predicted value will fall within that specified range.

C

• • •

CONFUSION MATRIX

A confusion matrix is a NxN matrix that depicts 'how successful a classification model's predictions were.' one axis of the matrix represents the label that the model predicted, and the other axis depicts the actual label.

The confusion matrix, in case of a multi-class model, helps in determining the mistake patterns. Such a confusion matrix contains sufficient information to calculate performance metrics, like 'precision' and 'recall.'

CONTINUOUS VARIABLE

A continuous variable can have an infinite number of values within a particular range. Its nature contrasts with the 'discrete variables' or 'discrete feature.' For example, if you can express a value as a decimal number, then it is a continuous variable.

CONVERGENCE

In simple words, convergence is a point where additional training on the data will not improve the model anymore. At convergence, the 'training loss' and 'validation loss' do not change with further iterations.

In deep learning models, loss values can stay constant or unchanging for many numbers of iterations, before finally descending. This observation might produce a false sense of convergence.

CONVEX FUNCTION

A convex function is usually a 'U-shaped' curve. In degenerate cases, however, the convex function is shaped like a line. These functions represent loss functions. The sum of two convex functions is always a convex function. Example of a convex function is 'Log Loss' function.

C

...

CORRELATION

Correlation is the measure of 'how closely the two data sets are correlated.' Take for example two data sets, 'subscriptions' and 'magazine ads.' When more ads get displayed. More subscriptions for a magazine get added, i.e., these data sets correlate. A correlation coefficient of '1' is a perfect correlation, 0.8 represents a strong correlation while a value of 0.12 represents weak correlation.

The correlation coefficient can also be negative. In the cases where data sets are inversely related to each other, a negative correlation might occur. For example, when 'mileage' goes up, the 'fuel costs' go down. A correlation coefficient of -1 is a perfect negative correlation.

COVARIANCE

Covariance is the 'measure of association between the average value of two variables, diminished by the product of their average values.' It represents how 'two variables vary together from their mean.'

CROSS-ENTROPY

Cross-entropy is a means to quantify the difference between two probability distributions. It is a generalization of 'Log Loss' function to multi-class classification problems.

D

...

DATA-DRIVEN DOCUMENTS OR D3

D3 is a popular JavaScript library used by the data scientists, to present their results of the analysis in the form of interactive visualizations embedded in web pages.

DATA MINING

Data mining is the analysis of large structured datasets by a computer to find hidden patterns, relations, trends and insights within it. Data mining comes from data science.

DATASET

A data set is a collection of structured information, used as 'examples' in machine learning.

DATA SCIENCE

Data science is the field of study employing scientific methods, processes, and systems to extract knowledge and insights from complex data in various forms.

DATA STRUCTURE

A data structure represents the way in which the information of the data is arranged. Example, array structure or 'tree' data structure.

DATA WRANGLING

Data wrangling or data munging is the conversion of data to make it easier to work with. It is achieved by using scripting languages like 'Perl.'

D

...

DECISION BOUNDARY

Decision boundary is the separating line between the classes learned by a model in a 'binary class' or 'multiclass classification' problems.

DECISION TREES

A decision tree represents the number of possible decision paths and an outcome for each path, in the form of a tree structure.

DEEP LEARNING OR DEEP MODEL

It is a type of 'neural network' containing a multi-level algorithm to process data at increasing level of abstraction. For example, the first level of the algorithm may identify lines, and the second recognizes the combination of lines as shapes and the third level recognizes the combination of shapes as objects.

Deep models depend on 'trainable nonlinearities.' It is a popular model for image classification.

DENSE FEATURE

It is a function feature in which most values are non-zero. A dense feature is typically a 'Tensor' of floating point values.

DEPENDENT VARIABLE

A dependent variable's value is influenced by the value of an independent variable. For example, 'The magazine ad budget' is an independent variable value. However, the number of 'subscriptions' made is dependent on the former variable.

D

...

DIMENSION REDUCTION

Dimension reduction is the extraction of one or more ‘dimensions’ that ‘capture’ as many variations in the data as possible. It is implemented with a technique called ‘Principal component analysis.’ Data reduction is useful in finding a small subset of data that captures ‘most of the variation’ in a given dataset.

DISCRETE FEATURE OR DISCRETE VARIABLE

A discrete feature is a variable whose possible values are finite. It contrasts with ‘continuous feature.’

DROPOUT REGULARIZATION

A dropout regularization ‘removes a random selection of a fixed number of units in a network layer’ for a single gradient step. This form of regularization is used in training neural networks. The more the number of units dropped out, the stronger will be the regularization.

DYNAMIC MODEL

A dynamic model is trained online with the continuously updated data. In such a model, the data keeps entering it continually.

E

• • •

EARLY STOPPING

If the loss on 'validation dataset' increases, the 'generalization performance' worsens. Hence, the model training has to be ended. It is known as early stopping. 'Early stopping' is a method of regularization in which the model training ends before 'training loss' finish decreasing. If the loss on 'validation dataset' increases, the 'generalization performance' worsens. Hence, the model training has to be ended. It is known as early stopping. 'Early stopping' is a method of regularization in which the model training ends before 'training loss' finish decreasing.

EMBEDDINGS

Embeddings are categorical features represented as continuous-valued features. An embedding is a translation of a 'High-dimensional vector' into a 'Low dimensional space.'

Embeddings are trained by 'Backpropagating loss' like any other parameter in a neural network.

EMPIRICAL RISK MINIMIZATION (ERM)

ERM is the selection of 'model function' that minimizes training losses. It contrasts with 'Structural risk minimization.'

ENSEMBLE

To 'ensemble' is to merge the predictions of multiple models. For example, 'deep and wide models' are the ensemble. An ensemble can be created via different initializations, different overall structures or different hyperparameters.

E, F

...

ESTIMATOR

An estimator encapsulates or contains the logic that builds a TensorFlow graph and runs a TensorFlow session.

EXAMPLE

An 'example' represents 'one row' of a given data set. It also contains one or more 'features.' It might also carry labels. Hence, examples can be labeled or unlabeled.



FALSE NEGATIVE OR FN

If a model mistakenly predicts an example to be of 'negative class,' the outcome is called false negative. Example, if a model predicts an email as 'not spam'(negative class) but it actually was 'spam.'

FALSE POSITIVE OR FP

If a model mistakenly predicts an example to be of 'positive class,' the outcome is called false positive. Example, if a model predicts an email as 'spam'(positive class) but it actually was 'not spam.'

FALSE POSITIVE RATE

Mathematically, the false positive rate is defined as;

$\text{FP rate} = (\text{Number of false positives}) / (\text{Number of false positives} + \text{number of true negatives})$

FP rate is represented by x-axis in a ROC curve.

F

...

FEATURE

A feature is an input variable value used to make predictions. It represents 'pieces of measurable information' about something. For example, a person's age, height, and weight represent three features about him/her. A feature can also be called property or an attribute.

FEATURE COLUMNS OR FEATURECOLUMNS

A feature column is a set of related features of an example. For instance, 'a set of all possible languages,' a person might know, will be listed under one feature column. A feature column might contain a single feature as well.

FEATURE CROSS

A feature cross represents non-linear relationships between features. It is formed by multiplying or taking a Cartesian product of individual features.

FEATURE ENGINEERING

Feature engineering involves 'determining which feature will be useful in training a model.' The 'raw data' from log files and other sources is then converted into the said features. Feature engineering is also referred to as 'feature extraction.'

FEATURE SET

It is the 'set of features' on which a machine learning model trains. Take, for example, the model of a used car, its age, distance covered, etc. These 'set of features' can be used to predict the price of that car.

G

...

GATE OR GENERAL ARCHITECTURE FOR TEXT ENGINEERING

GATE is an open-source Java-based framework for natural language processing tasks. This framework allows the user to integrate other tools designed to be plugged into it.

GENERALIZATION

Generalization is the ability of a model to judge correct predictions based on 'fresh and unseen' data, and not on the data previously used to train the model.

GENERALIZED LINEAR MODEL

A generalized linear model is a generalization of 'least squares regression models' based on Gaussian noise, to other types of models based on other types of noises. The examples of generalized linear models are 'Logistic regression' and 'multiclass regression.'

A generalized linear model cannot learn 'new features' like a deep learning model does.

GRADIENT

A gradient represents the 'vector of partial derivatives' concerning to all the independent variables. A gradient always points towards the 'steepest ascent.'

GRADIENT CLIPPING

Gradient clipping is the method of ensuring numerical stability by 'capping' gradient values before applying them.

G

...

GRADIENT BOOSTING

Gradient boosting produces a prediction model in the form of an ensemble of weak prediction models. This is a machine learning technique for regression and solving classification problems.

Gradient boosting builds the model stage-wise and generalizes them by allowing optimization of arbitrary differentiable loss functions.

GRADIENT DESCENT

Gradient descent is a loss minimization technique, which involves computing of gradients of loss with respect to the model's parameters, learned or trained on training data. Gradient descent works by adjusting parameters and finding the optimum combination of 'weights' and bias to minimize loss.

GRAPH

A graph represents a 'computation specification' to be processed in TensorFlow. Such a graph is visualized using TensorBoard. The nodes on the graph depict operations and edges represent the passing of the result as an operand to another operation (or Tensor).

H

• • •

~~HEURISTIC~~

A heuristic is a practical solution to a problem that aids in learning and making progress.

~~HIDDEN LAYER~~

A hidden layer in a neural network lies between the input layer (or feature) and the output layer (or prediction). A neural network can contain single or multiple hidden layers.

~~HINGE LOSS~~

A hinge loss is a loss function designed for classification models, to find the decision boundary as far as possible from each training example. A hinge loss function maximizes the margin between examples and the boundary.

~~HISTOGRAM~~

A histogram represents the distribution of numerical data through a vertical bar graph.

~~HOLDOUT DATA~~

These are the datasets that are intentionally held-out during the model's training. Holdout data helps in evaluation of the model's ability to generalize to data, other than the data it was trained on. Examples of holdout datasets are validation dataset and test data set.

~~HYPERPARAMETER~~

The parameters that can be 'changed' or 'tweaked' during successive training runs of a model are known as hyperparameters.



•••

INDEPENDENTLY AND IDENTICALLY DISTRIBUTED (IID)

IID represents a collective of data or variables that have 'same probability distribution' as the others and are mutually independent. In case of IIDs, the probability of a predicted outcome is 'no more' or 'less' likely than any other prediction.

Example of an IID is 'a fairly rolled dice.' Here, all the faces always have an equal probability of coming up, irrespective of the number of times the number faces that already came up.

INFERENCE

The inference is the process through which a trained model makes predictions to unlabeled examples. This definition is in regards to machine learning.

INPUT LAYER

The input layer is the first layer to receive the input data in a neural network.

INTER-RATER AGREEMENT

Inter-rater agreement is a way to measure the 'agreement' between human raters while undertaking a task. A disagreement amongst the raters calls for the improvement in 'task instructions.'

K, L

...

KERNEL SUPPORT VECTOR MACHINES (KSVMS)

A K SVM maps the input data vectors to a higher dimensional space for maximizing the margin between positive and negative classes. K SVMs employ hinge loss as a loss function.

K-MEANS CLUSTERING

It is a data-mining algorithm to classify or group or ‘cluster’ ‘N’ number of objects based on their features into ‘K’ number of groups (or clusters).

K-NEAREST NEIGHBORS OR KNN

It is a machine learning algorithm that examines ‘k’ number of ‘neighbors’ to classify things based on their similarity. Here, ‘similarity’ means the comparison of ‘feature values’ in the neighbors being compared.

LATENT VARIABLE

Latent variables are hidden variables, whose presence is inferred by directly measuring the observed variables. The inference of these variables is made through a mathematical model.

LABEL

In machine learning terms, a label represents the ‘answer’ or ‘result’ associated with an example.



• • •

LAYER

A layer is a set of neurons that process a set of input features or the output of those neurons in a neural network.

LIFT

Life signifies 'how frequently a pattern will be observed by chance.' if the lift is 1, then the pattern is supposed to be occurring coincidentally. The higher the lift, the higher is the chance that the occurring pattern is real.

LINEAR REGRESSION

Linear regression is the method of graphically expressing the relationship between a scalar dependent variable 'y,' and one or more independent variable 'X.' For example, the relationship between 'price' and 'sales' can be expressed with an equation as a straight line on the graph.

LOGISTIC REGRESSION

Logistic regression is a model similar to linear regression, and only the output result is made to fit the logistic function. In other words, the potential results are not continuous but 'specific set of categories.'

M

• • •

MACHINE LEARNING

Machine learning or ML involves the development of algorithms to figure out insights from extensive and vast data. 'Learning' refers to 'refining' of the models by supplying additional data, to make it perform better with each iteration.

MARKOV CHAIN

Markov chain is an algorithm, used to determine the possibility of occurrence of an event, based on which other events have already occurred. This algorithm works with the data of 'series of events.'

MATRIX

Matrix is merely a set of data arranged in rows and columns.

MEAN

Mean, or arithmetic mean is the average value of numbers.

MEAN ABSOLUTE ERROR

Mean Absolute error or MAE is the average error of predicted values as compared to the observed values.

M

• • •

MEAN SQUARED ERROR OR MSE

MSE is the average of the squares of all the predicted values as compared to the observed values.

MEDIAN

The central or middle value of a sorted data is called the median. If the number of values in data is even, the average value of the two central digits become the median.

MODE

For a given set of data values, the value that appears most frequently is called the mode. Mode, like median, is a way to measure the central tendency.

MODEL

In statistical analysis, modeling refers to the specification of a probabilistic relationship existing between different variables. A 'model' runs on algorithms and training data to 'learn' and then make predictions.

MONTE CARLO METHOD

Monte Carlo method is a technique to solve numerical problems by studying numerous randomly generated numbers, to find an approximate solution. Such a numerical problem is often challenging to solve by other mathematical methods.

Monte Carlo method is often used by Markov chain algorithm.

M, N

...

MOVING AVERAGE

Moving average represents the ‘continuous average’ of new time series data. The mean of such data is calculated at equal time intervals and is updated according to the most recent value, while the older value gets dropped.

MULTIVARIATE ANALYSIS

The analysis of ‘dependency of multiple variables over each other’ is called the multivariate analysis.



N-GRAM

N-gram is the ‘scanning of patterns in a sequence of ‘N’ items.’ It is typically used in natural language processing. For example, unigram analysis, bigram analysis, trigram analysis and so on.

NAIVE BAYES CLASSIFIER

A naive Bayes classifier is an algorithm based on Bayes’ theorem, which classifies features with an assumption that ‘every feature is independent of every other feature.’ This classification algorithm is called ‘naive’ because all the features might not necessarily be independent, and it becomes one downside of this algorithm.

NATURAL LANGUAGE PROCESSING

Natural language processing or NLP is a collection of techniques to structurize and process raw text from human spoken languages to extract information.

N

...

NEURAL NETWORK

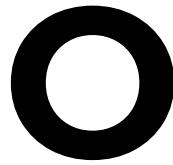
A neural network uses algorithmic processes that mimic the human brain. It attempts to find insights and hidden patterns from vast data sets. A neural network runs on learning architectures and is 'trained' on large data sets to make such predictions.

NORMAL DISTRIBUTION

Normal distribution or 'bell curve' or 'Gaussian distribution' is a continuous bell-shaped graph with the mean value at the center. It is a widely used distribution curve in statistics.

NULL HYPOTHESIS

A null hypothesis depicts that 'a single variable is not different from its mean or no variation exists between variables.' Hence, according to a null hypothesis, the given observations hold no 'statistical significance.'



• • •

OBJECTIVE FUNCTION

An objective function maximizes or minimizes a ‘result’ (or objective) by changing the values of other quantities like decision variables, constraints and the result into an objective function.

ONE HOT ENCODING

One hot encoding converts categorical variables into numerical, to make it interpretable to the learning model.

ORDINAL VARIABLE

Ordinal variables are ordered variables with discrete values.

OUTLIER

Observations that diverge far away from the overall pattern in a sample are called outliers. An outlier may also indicate an error or rare events.

OVERFITTING

An overly complicated model of data that takes too many outliers or ‘intrinsic data quirks’ into account. Overfitting model of training data is not much useful in finding patterns in test data.

P

• • •

P VALUE

P value depicts the probability of getting a result equal to or more than the actual observation, under the null hypothesis. It is a measure of 'the gap shown between the groups when there actually isn't any gap.'

PERCEPTRON

Perceptron is the simplest neural network, in which a single neuron approximates 'n' binary inputs.

PIVOT TABLE

A pivot table allows for easily rearranging long lists of data and summarize them. The act of rearranging the data is known as 'pivoting.' Pivot table also allows for the dynamic rearrangement of the data by just creating a pivot summary. It takes away the need for employing a formula or copying to data arrangement.

POISSON DISTRIBUTION

It is the distribution of independent events over a defined time period and space. Poisson distribution is used to predict the probability of occurrence of an event.

PREDICTIVE ANALYTICS

Predictive analytics involves extraction of information from existing data sets to determine patterns and insights. These patterns and insights are used to predict future outcomes or event occurrences.

P

• • •

PRECISION AND RECALL

Precision is simply the measure of 'true positive predictions' out of all the positive predictions.

Mathematically, 'Precision'=(True positive predictions)/(True positives+false positives).

Recall, on the other hand, is the measure of 'number of correct positive predictions.'

For example, take a visual recognition model that recognizes 'oranges.' It recognizes seven oranges in a picture containing ten oranges with some apples.

Out of those seven oranges, five are actually oranges (true positives), and the rest two are apples (false positives).

Then, 'precision'=5/7 and 'recall'=5/10.

PREDICTOR VARIABLES

Predictor variables make predictions for dependent variables.

PRINCIPAL COMPONENT ANALYSIS

This algorithm analysis highest variance in the given data. This variance value is tagged as the principal component.

PRIOR DISTRIBUTION

In Bayesian statistics, 'prior' probability distribution of an uncertain quantity is based on assumptions and beliefs, without taking any evidence into account.

Q, R

...

QUANTILES AND QUARTILES

Division of sorted values into groups having the same number of values is called a 'quantile' group. If the number of these groups is four, they are called 'quartiles.'

R

R is an open-source programming language for statistical analysis and graph generation, available for different operating systems.

RANDOM FOREST

An algorithm that employs 'a collection of tree data structures' for the classification task. The input is classified or 'voted' for by each tree. 'Random forest' chooses the classification with the highest 'votes' compared to all the trees.

RANGE

The range is the difference between the highest and the lowest value in a given set of numbers. For example, consider the set 2,4,5,7,8,9,12. The range=12-2 i.e. 10.

REGRESSION

Regression aims to measure the dependency of one dependent variable and other changing variables. Examples, linear regression, logistic regression, lasso regression, etc.

R

...

REINFORCEMENT LEARNING

Reinforcement learning or RL is a learning algorithm that allows a model to interact with an environment and make decisions. The model is not given specific goals, but when it does something 'right', it is given feedback. This 'reinforcement' helps the classification model in learning to make right predictions. RL model also learns from its 'past' experiences.

RESPONSE VARIABLE

The response variable is the one that can be manipulated by other variables. It is also called dependent variable.

RIDGE REGRESSION

Ridge regression performs the 'L2 regularization' function on the optimization objective. In other words, it adds the factor of the sum of squares of coefficients to the objective.

ROOT MEAN SQUARED ERROR OR RMSE

RMSE denotes the standard deviation of prediction errors from the regression line. It is simply the square root of the mean squared error. RMSE signifies the 'spread' or 'concentration' of data around the regression line.

S

• • •

S CURVE

As the name suggests, 'S-curve' is a graph shaped like the letter 'S.' It is a curve that plots variables like cost, number, population, etc. against time.

SCALAR

A scalar quantity represents the 'magnitude' or 'intensity' of a measure and not its direction in space or time. For example, temperature, volume, etc.

SEMI-SUPERVISED LEARNING

Semi-supervised learning involves the use of extensive 'unlabeled data' at the input. Only a small data is 'labeled' for the model to 'learn' to make the right classifications without much external supervision.

SERIAL CORRELATION

Serial correlation or autocorrelation is a pattern in a series, where each value is directly influenced by the value next to it or preceding it. It is calculated by shifting a time-series over the numerical series by an interval called 'lag.'

SKEWNESS

Skewness represents symmetry of distribution or a data set, to the left or the right from its center point.

S

...

SPATIOTEMPORAL DATA

A spatiotemporal data includes the space and time information about its values. In other words, it is a time-series data with geographic identifiers.

STANDARD DEVIATION

Standard deviation represents the 'dispersion of the data.' It is the square root of the variance to show how far an observation is from the mean value.

STANDARD ERROR

Standard error signifies the 'statistical accuracy of an estimate.' It is equal to the standard deviation of the sampling distribution of a statistic.

STANDARD NORMAL DISTRIBUTION

It is same as the normal distribution, just with a mean of '0' and standard deviation equal to '1.'

STANDARDIZED SCORE

Standard score, normal score or Z-score is the 'transformation of the raw score for evaluating it in reference with the standard normal distribution, by converting it into units of standard deviation above or below the mean.'

S

...

STRATA

Division of the data into homogeneous groups and drawing random samples from each group represents a 'strata.' for example, forming 'strata' of the population or demographic data.

SUPERVISED LEARNING

Supervised learning involves using algorithms to classify the input into specific pre-determined or known classes. In such a case, the prediction made by the model is based on a 'given set of predictors.'

Some examples of supervised learning algorithms are Random forest, decision tree, and kNN, etc.

SUPPORT VECTOR MACHINE OR SVM

A support vector machine is a discriminative classifier, which plots data-items in 'n' dimensional space. Here, 'n' represents the number of features each data-item (or data point) has. The data points are plotted on the coordinates (support vectors).

T,U

...

T-DISTRIBUTION

T-distribution is the 'sampling of all the possible values instead of actually using them' on the normal distribution curve. It is also known as 'Student's T-distribution.'

TYPE I ERROR

The incorrect decision to reject the null-hypothesis is called type I error.

TYPE II ERROR

The incorrect decision to retain or keep the null-hypothesis is called type II error.

T-TEST

It is the analysis of 'two population datasets' by finding the difference of their populations.



UNIVARIATE ANALYSIS

The univariate analysis' purpose is to describe the data. It analyzes the dependency of a single predictor and the response variable.

UNSUPERVISED LEARNING

An algorithm that classifies groups of data without knowing what the groups will be. There is no target or outcome variable to predict or estimate. Unsupervised learning focuses majorly on learning from the underlying data based on its attributes.

V,W

•••

VARIANCE

It is the ‘variation’ of the numbers in a given data from the mean value. Variance represents the magnitude of differences in a given set of numbers.

VECTOR

In mathematical terms, vector denotes the quantities with magnitude and direction in the space or time. In data science terms, it means ‘ordered set of real numbers, each representing a distance on a coordinate axis.’ For example, velocity, momentum, or any other series of details around which the model is being built.

VECTOR-SPACE

Vector space is the collection of vectors. For example, a matrix is a vector-space.

WEKA

Weka is a collection of machine learning algorithms and tools for mining data. Using Weka, the data can be pre-processed, regressed, classified, associated with rules and visualized.

INTERNATIONAL SCHOOL OF
ARTIFICIAL INTELLIGENCE & DATA SCIENCE

6 MONTH CERTIFICATE IN DATA SCIENCE & MACHINE LEARNING

6 months | Online | 6 Terms

• • •

Accelerate your career with Data Science &
Machine Learning & become a data leader of tomorrow

APPLY NOW



Contact us

www.insaid.co/contact

Email: info@insaid.co