# Passcert

Higher Quality, better service!

# Q&A

*Http://www.passcert.com*

*We offer free update service for one year.*

**Exam** : **DSA-C02**

**Title** : SnowPro Advanced: Data Scientist Certification Exam

**Version** : DEMO

1.Which type of Machine learning Data Scientist generally used for solving classification and regression problems?
A. Supervised
B. Unsupervised
C. Reinforcement Learning
C. Instructor Learning
D. Regression Learning
**Answer:** A
**Explanation:**
Supervised Learning
Overview:
Supervised learning is a type of machine learning that uses labeled data to train machine learning models. In labeled data, the output is already known. The model just needs to map the inputs to the respective outputs.
Algorithms:
Some of the most popularly used supervised learning algorithms are:
· Linear Regression
· Logistic Regression
· Support Vector Machine
· K Nearest Neighbor
· Decision Tree
· Random Forest
· Naive Bayes
Working:
Supervised learning algorithms take labelled inputs and map them to the known outputs, which means you already know the target variable.
Supervised Learning methods need external supervision to train machine learning models. Hence, the name supervised. They need guidance and additional information to return the desired result.
Applications:
Supervised learning algorithms are generally used for solving classification and regression problems. Few of the top supervised learning applications are weather prediction, sales forecasting, stock price analysis.

2.Which of the learning methodology applies conditional probability of all the variables with respective the dependent variable?
A. Reinforcement learning
B. Unsupervised learning
C. Artificial learning
D. Supervised learning
**Answer:** D
**Explanation:**
Supervised learning is a type of machine learning where we train a model using labeled data. In this learning paradigm, the model learns from the training data to make predictions or infer mappings. Conditional probability often plays a role in this, especially in algorithms like Naive Bayes, where the goal

is to compute the probability of a certain class given the features (variables) of the data, which is fundamentally a conditional probability.
Here's a brief rundown of the other options:
A. Reinforcement learning: This is about agents who take actions in an environment to maximize cumulative reward. It's not centered around conditional probability of variables.
B. Unsupervised learning: This is about finding patterns in data without labeled responses. Methods such as clustering and dimensionality reduction fall into this category.
C. Artificial learning: This is not a standard term in the field of machine learning or data science.

3.In a simple linear regression model (One independent variable), If we change the input variable by 1 unit.
How much output variable will change?
A. by 1
B. no change
C. by intercept
D. by its slope
**Answer:** D
**Explanation:**
What is linear regression?
Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.
Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be an explanatory variable, and the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.
A linear regression line has an equation of the form Y = a + bX, where X is the explanatory variable and Y is the dependent variable. The slope of the line is b, and a is the intercept (the value of y when x = 0).
For linear regression Y=a+bx+error.
If neglect error then Y=a+bx. If x increases by 1, then Y = a+b(x+1) which implies Y=a+bx+b. So Y increases by its slope.
For linear regression Y=a+bx+error. If neglect error then Y=a+bx. If x increases by 1, then Y = a+b(x+1) which implies Y=a+bx+b. So Y increases by its slope.

4.There are a couple of different types of classification tasks in machine learning, Choose the Correct Classification which best categorized the below Application Tasks in Machine learning?
· To detect whether email is spam or not
· To determine whether or not a patient has a certain disease in medicine.
· To determine whether or not quality specifications were met when it comes to QA (Quality Assurance).
A. Multi-Label Classification
B. Multi-Class Classification
C. Binary Classification
D. Logistic Regression
**Answer:** C

**Explanation:**
The Supervised Machine Learning algorithm can be broadly classified into Regression and Classification Algorithms. In Regression algorithms, we have predicted the output for continuous values, but to predict the categorical values, we need Classification algorithms.
What is the Classification Algorithm?
The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data. In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups. Such as, Yes or No, 0 or 1, Spam or Not Spam, cat or dog, etc. Classes can be called as targets/labels or categories. Unlike regression, the output variable of Classification is a category, not a value, such as "Green or Blue", "fruit or animal", etc. Since the Classification algorithm is a Supervised learning technique, hence it takes labeled input data, which means it contains input with the corresponding output. In classification algorithm, a discrete output function(y) is mapped to input variable(x).
y=f(x), where y = categorical output
The best example of an ML classification algorithm is Email Spam Detector.
The main goal of the Classification algorithm is to identify the category of a given dataset, and these algorithms are mainly used to predict the output for the categorical data.
The algorithm which implements the classification on a dataset is known as a classifier. There are two types of Classifications:
Binary Classifier: If the classification problem has only two possible outcomes, then it is called as Binary Classifier.
Examples: YES or NO, MALE or FEMALE, SPAM or NOT SPAM, CAT or DOG, etc. Multi-class Classifier: If a classification problem has more than two outcomes, then it is called as Multi-class Classifier.
Example: Classifications of types of crops, Classification of types of music.
Binary classification in deep learning refers to the type of classification where we have two class labels – one normal and one abnormal.
Some examples of binary classification use:
· To detect whether email is spam or not
· To determine whether or not a patient has a certain disease in medicine.
· To determine whether or not quality specifications were met when it comes to QA (Quality Assurance).
For example, the normal class label would be that a patient has the disease, and the abnormal class label would be that they do not, or vice-versa.
As is with every other type of classification, it is only as good as the binary classification dataset that it has – or, in other words, the more training and data it has, the better it is.

5.Which of the following method is used for multiclass classification?
A. one vs rest
B. loocv
C. all vs one
D. one vs another
**Answer:** A
**Explanation:**
Binary vs. Multi-Class Classification

Classification problems are common in machine learning. In most cases, developers prefer using a supervised machine-learning approach to predict class tables for a given dataset. Unlike regression, classification involves designing the classifier model and training it to input and categorize the test dataset. For that, you can divide the dataset into either binary or multi-class modules.

As the name suggests, binary classification involves solving a problem with only two class labels. This makes it easy to filter the data, apply classification algorithms, and train the model to predict outcomes. On the other hand, multi-class classification is applicable when there are more than two class labels in the input train data. The technique enables developers to categorize the test data into multiple binary class labels.

That said, while binary classification requires only one classifier model, the one used in the multi-class approach depends on the classification technique. Below are the two models of the multi-class classification algorithm.

One-Vs-Rest Classification Model for Multi-Class Classification

Also known as one-vs-all, the one-vs-rest model is a defined heuristic method that leverages a binary classification algorithm for multi-class classifications. The technique involves splitting a multi-class dataset into multiple sets of binary problems. Following this, a binary classifier is trained to handle each binary classification model with the most confident one making predictions.

For instance, with a multi-class classification problem with red, green, and blue datasets, binary classification can be categorized as follows:

Problem one: red vs. green/blue

Problem two: blue vs. green/red

Problem three: green vs. blue/red

The only challenge of using this model is that you should create a model for every class. The three classes require three models from the above datasets, which can be challenging for large sets of data with million rows, slow models, such as neural networks and datasets with a significant number of classes.

The one-vs-rest approach requires individual models to prognosticate the probability-like score. The class index with the largest score is then used to predict a class. As such, it is commonly used for classification algorithms that can naturally predict scores or numerical class membership such as perceptron and logistic regression.