Sub :- ML (Machine learning)

18/11/2021

Assignment

Name :- Aditya Sunil Chandrikapure
ID :- 19007002

Que1 What are different Types of Machine learning algorithms?

⇒ Types of Machine learning algorithms :-

→ There are four types of machine learning algorithms as follows.

(i) Supervised
(ii) Semi - supervised
(iii) Unsupervised
(iv) Reinforcement

(1) Supervised :-

→ Supervised learning algorithms are the ones that involve direct supervision (we the title) of the operation.

→ In this case, the developer labels sample data corpus and set strict boundaries upon which the algorithm operates.

→ The primary purpose of supervised learning is to scale the scope of data and to make predictions of unavailable, future or unseen data based on labeled sample data.

→ supervised machine learning includes two major processes.

→ classification
→ Regression.

→ Classification :-

→ It is the process where incoming data is labeled based on the past data samples and manually trains the algorithm to recognize certain types of object and categorize them accordingly.

→ Regression :-

→ It is the process of identifying patterns and calculating the prediction of continuous outcomes.

→ The system has to understand the numbers, their values etc.

The most widely used supervised algorithms are.

→ Linear Regression
→ Logistical Regression
→ Random forest.
→ Gradient Boosted Trees.
→ Support vector machines (SVM)
→ Neural Networks
→ Decision Trees.
→ Nearest Neighbour.

(11) Semi-Supervised Machine learning algorithm :-

→ semi-supervised learning algorithms represent a middle ground between supervised and unsupervised algorithms.

→ In essence, the semi-supervised model combines some aspects of both into a thing of its own.

* How semi-supervised algorithm work.

(1) A semi-supervised machine learning algorithm uses a limited set of labeled sample data to shape the requirement of the operation

(11) The limitation result in a partially trained model that later gets the task to label the unlabeled data.

(111) Due to the limitation of the sample data set, the result are considered pseudo-labeled data.

(IV) finally, labeled and pseudo-labeled data sets, are combined. Which creates a distinct algorithms that combine descriptive and predictive aspect of supervised and unsupervised learning).

(iii) Unsupervised Machine learning Algorithms :-

→ Unsupervised learning is one that does not involve direct control of the developer.

→ If the main point of supervised machine learning that we know the result and need to sort out the data, then in the case of unsupervised machine learning algorithm the desired result are unknown and yet to be defined.

→ The unsupervised learning feeds on the unlabled data.

→ Unsupervised learning algorithms apply the following technique to describe the data.

(i) Clustring :-

→ It is an exploration of data used to segment it into meaningful groups based on their internal patterns without priors knowledge of group credentials.

(ii) Dimensionality Reduction :-

→ There is a lot of noise in the incoming data.

→ machine learning algorithms used dimensionality reduction to remove this noise while distilling the relevent information

(iv) Reinforcement Machine Learning Algorithms :-

→ Reinforcement learning represents what is commonly understood as machine learning artificial intelligence.

→ Reinforcement learning is all about developing a self-supervised system that, throughout contiguous sequences of tries and fails, improves itself based on the combination of labeled data and interactions with incoming data.

→ Reinforcement ML uses the techniques called exploration / exploitation.

→ The mechanics are simple - the action takes place, the consequences are observed, and the next action considers the results of the first action.

→ In the center of reinforcement learning algorithm are reward signals that occur upon performing specific tasks.

Que2 What is the difference between training and testing in machine learning?

⇒

Training	Testing
→ The training data set is the general term for the samples used to create the model	→ test or validation data set is used to quality performance.
→ we use it to fit the model	→ we used it to test the model.
→ Training set is implement to build up a model.	→ A test is to validate the model built.
→ Here, we have the complete training data set, we can extract feature and train to fit a model and so on.	→ Here, once the model is obtained, you can predict using the model obtained on the training set.
→ Training are necessary to teach the algorithm how to make accurate predictions in accordance with the goals of an AI project.	→ testing are necessary to developed the data set for accurate predictions in accordance with the goals of an AI project.

Q.3 Why we split data into train and test while constructing a ML model?

⇒ → The Reason is that when the dataset is split into train and test sets, there will not be enough data in the training data for the model to Learn an effective mapping inputs and outputs.

→ There will also not be enough data in the test set to effectively evaluate the model performance.

→ The estimated performance could be overly optimistic or overly pessimistic.

Q.4 What are the three stages to build the hypotheses or model in machine learning? explain with the help of example.

⇒ The Three stages to build the hypotheses or model in machine learning as follows.

(i) Model building
(ii) Model testing
(iii) Model Applying

(i) Model building :-

→ Choose a suitable algorithm for the model and train it according to the requirements.

→ It's one of the critical step of model in ML involving correct choise of algorithms.

(ii) Model Training :-

→ In this step we focus on the training dataset and map the accuracy of the model through test dataset.

(iii) Applying the model :-

→ Make the requirment change after testing and use the final model

for real time projects.

Example :-

→ An example of model building is approximation the target function and performs mappings of inputs and outputs is called a hypothesis in machine learning.

→ The training dataset is used to learn a hypothesis and test dataset is used to evaluate it.

Q.5 State the key difference between AI and ML.

Artificial Intelligence	Machine learning
→ Artificial intelligence is a technology which enables a machine to simulate human behavior.	→ Machine learning is a subset of AI which allows a machine to automatically learn from past data without programming explicitly.
→ The goal of AI is to make a smart computer system like humans to solve complex problems.	→ The goal of ML is to allow machine to learn from data so that they can give accurate output.
→ In AI, we make intelligent systems to perform any task like human	→ In ML, we teach machine with data to perform a particular task and give an accurate result.
→ Machine learning and deep learning are the two main subsets of AI.	→ Deep learning is a main subset of machine learning.
→ AI has a very wide range of scope	→ machine learning has a limited scope.

→ AI is working to create an intelligent system which can perform various complex task.	→ Machine learning is working to create machine that can perform only those specific tasks for which they are trained.
→ The main applications of AI are siri, customer support using catbots, expert system, online game playing, intelligent humanoid robot etc.	→ The main applications of machine learning are Online recommender system, Google search algorithms, facebook auto friend tagging suggestions etc.
→ On the basis of capabilities, AI can be divided into three types. which are, Weak AI, General AI, and strong AI.	→ Machine learning can also be divided into mainly three types that are supervised learning, Unsupervised learning, and Reinforcement learning.

Q.6 What is the difference between supervised and unsupervised learning?

Supervised learning	Unsupervised Learning
→ Supervised learning algorithms are trained using labeled data.	→ Unsupervised learning algorithms are trained using unlabelled data.
→ Supervised learning model takes direct feedback to check if it is predicting correct output or not	→ Unsupervised learning model does not take any feedback.
→ supervised learning model predicts the output.	→ Unsupervised learning model finds the hidden patterns in data.
→ In supervised learning, input data is provided to the model along with the output.	→ In unsupervised learning, only input data is provided to the model.
→ The goal of supervised learning is to trained the model so that it can predict the output when it is given new data.	→ The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset.

→ supervised learning needs supervision to train the model	→ unsupervised learning does not need any supervision to train the model.
→ Supervised learning can be categorized in Classification and Regression problems.	→ Unsupervised learning can be classified in Clustring and Associations problems.

Q.7 Explain the difference between classification and Regression?

Regression	classification.
→ In Regression, the output variable must be of continuous nature or real value.	→ In classification, the output variable must be a discrete value.
→ The task of the regression algorithm is to map the input value (x) with the continuous output variable (y).	→ The task of the classification algorithm is to map the input value (x) with the discrete output variable (y).
→ Regression Algorithm are used with continuous data.	→ classification algorithm are used with discrete data.
→ In Regression, we try to find the best fit line, which can predict the output more accurately.	→ In classification, we try to find the decision boundary, which can divide the dataset into different classes.
→ The regression algorithm can be further divide into linear and Non-linear Regression.	→ The classification algorithm can be divided into binary classifier and multi-class classifiers.

Q.8 Write a note on Geometric model, logical model, probabilistic model.

⇒ (1) Geometric model :-

→ In geometric model, features could be described as points in two dimensions (x-axis and y-axis) or three dimensional space. (x, y, and z).

→ Even when features are not physically geometric, they could be modelled in as geometric manner.

→ for example, temperature as a function of time can be modelled into two axis.

(2) Logical model :-

→ There are mainly two kinds of logical models - Tree models and Rule models.

→ Rule models consist of a collection of implication or IF-THEN rules, for ~~tree~~

→ For Tree based models, then IF-part defined the segment and the THEN part defines the behaviour of the model for this segment rule models follows the same reasonings.

(3) Probabilitic Model :-

→ Probabilistic model in machine learning is the use of codes of statistics to do the data examination.

→ probabilistic model are presented as a preventing idiom to to define the world.

→ those were described by using random variable for example building socks believed together. by probabilistic relationship.

Q.9 Explain the stages of building of machine learning model.

⇒ Stages of building machine learning model as follows.

(i) Data collection.
(ii) Data preparation
(iii) choose model
(iv) Train the model
(v) Evaluate the model.
(vi) Parameter training
(vii) make predictions

(i) Data collection :-

→ The quality & quality of your data dictate how accurate our model is the outcome of this step is generally a representation of data which we will use for training

→ Using pre-collected data, by way of datasets from kaggle etc, still fit into this step.

(ii) Data preparation :-

→ Wrangle data and prepare it for training.

→ Clean that which may required it.

→ Randomise data which erases the effects of particular order in which we collected and otherwise prepared our data.

→ visualize data. split into training and evaluations sets.

(iii) Choose Model :-

→ Different algorithms are for different tasks. choose right one.

(iv) Train the model :-

→ The goal of training is to answer a question or make a prediction correctly as often as possible.

→ Each iteration of process is a training step.

(v) Evaluate the model :-

→ uses some matric or combination of matrics to "measure" objective performance of model.

→ Test the model against perviously unseen data.

Q.10. What is the difference between feature selection and feature lea extraction?

Feature selection	Feature extraction.
→ Problem of selecting some subset of the learning algorithms input variables upon which it should focus attention, while ignoring the rest.	→ Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes.
→ Especially when dealing with a large number of variables there is a need for Dimensionality Reduction.	→ The general workflow involves applying feature extraction on given data to extract features and then apply feature selection with respect to the target variable to select a subset of data.
→ Feature selection can significantly improve a learning algorithm's performance.	→ The feature selection with respect to the target variable to select a subset of data. In effect, this helps improve the accuracy of a model.

→ many domains have tens to thousands of variable out of which most are irrelevant and redundant. It can significantly improve a learning algorithms performance.	→ Feature extraction can be used to extract the themes of a document collection, where doc- are represented by a set of key words and their frequency.

Q.11 Which neural network is best for binary classification?

⇒ ⊕ Neural network are a class of machine learning algorithms used to model complex patterns in different datasets using multiple hidden layers and a non-linear activation functions.

→ Neural network are trained iteratively using algorithm of optimization technique like gradient descent.

→ The use of single sigmoid & logistical neuron is the output layers is the main steps of binary classification neural Network.

→ This is because the output of a sigmoid / logistics function can be conviniently interrupted as the estimated probability that given input belongs to positive class.

Q.12 How are probability & statistics different?

⇒ → probability deals with predicting the likelihood of future events. while statics involves the analysis of frequent event of past events.

→ probability is primarily a theoratical branch of mathematical which studies the consequence of mathematical definitions.

→ statistics is primarily an applied branch of mathematics, which tries to make sence of observations in the real world.

Q.13 How to find RMSE & MSE?

⇒ → To complete RMSE, calculate the residual (difference between prediction & truth) for each data point.

→ Compute the norm of residual for each data point, compute the mean of residuals and take the square root of the mean.

→ RMSE is commonly used in supervised learning applications as RMSE uses and needs measurements at each predicated data point.

→ Root mean square error can be expressed as →

$$RMSE = \sqrt{\frac{\sum_{i}^{N} = {}^{''} \| y(i) - y^{\wedge}(i) \|^2}{N}}$$

where,

N → No, of data.

y(i) → the ith measurement

y'(i) → its corresponding predictions.

MSE (mean squared Error) :-

→ MSE is calculated by taking the average of the square of the different between the original and predicated value of the data.

$$MSE = 1/N \cdot \sum_{i=1}^{n} (\text{actual value} - \text{predicted values})^2$$

where,

N = total no. of observations froms in datasets.

Σ = denote that different between actual and predicted values taken en event.

: value ranging form 1 to n

Q.14 What are the important assumption of linear segression?

⇒ The important assumption of linear segression are as follows :-

→ Linear relationship between the features and target: Linear segression assumes the linear relationship between the dependent and independent variables.

→ Small or no. multicollinearity between the features: multicollinearity means high- correlation between the independent variables. Due to multicollinearity, it may difficult to find the true relationship betn the predictors and target variables.

→ Homoscedasticity Assumption :-
Homoscedasticity is a situation when the error term is the same for all the values of independent variables. With homoscedasticity, there should be no clear pattern distribution of data in the scatter plot.

→ Normal Distribution of error items :-
Linear regression assumes that the error terms should follow the normal distribution pattern. If error terms are not normally distributed. then confidence intervals will become either too wide or too narrow, which may cause difficult in finding coefficients.

It can be checked using the q-q plot. If the plot shows a straight line without any deviation which means the error is normally distributed.

→ No autocorrelations :-

The linear regression model assumes no autocorrelation in error terms. If there will be any correlation in the error term, it will drastically reduce the accuracy of the model. autocorrelation usually occurs if there is a dependency between residual errors.

Q.15 a) Need for polynomial Regression.
b) Equation of the polynomial Regression Model.
c) Steps for polynomial Regression.

⇒ a) Need for polynomial Regression :-

→ The goal of polynomial regression is to model a non-linear relationship between the independent and dependent variables. ~~some~~

→ some of these methods make use of a localized form of classical polynomial regression.

b) Equation of the polynomial Regression Model :-

→ Linear regression is polynomial regression of degree of and generally takes the form $y = mx + b$ where as the slope, and b is the y intercept. It could just as easily be written $y(x) = c_0 + c_1 x$ and c_1 being the slope and c_0 the y-intercept.

c) Steps for polynomial Regression :-

step1 :- Data pre-processing
step2 :- Linear Regression model
step3 :- polynomial Regression model