060010907-040250307 - Machine Learning

Unit-1

=====

List any four applications of machine learning and discuss how machine learning is useful in any two of the listed applications.

With appropriate example, explain supervised learning as machine learning task.

What is unsupervised learning? Explain with suitable example.

Compare supervised versus unsupervised machine learning. Discuss Similarities and differences.

What are the objectives of training and testing dataset? Explain cross validation with appropriate diagram.

Explain accuracy and recall as performance measures used to evaluate machine learning algorithms.

Define data preprocessing. Explain missing values and out-of-range values with respect to quality of data.

What is missing data in data preprocessing? Discuss how to handle missing data.

What is categorical data? Which are categorical data in following example? Explain, how encoding is done using LabelEncoder class in this example.

City	Age	Salary	Purchased
Bardoli	45	151000	Yes
Surat	51	164000	Yes
Valsad	42	134000	No
Surat	24	89000	Yes
Valsad	36		No

Compare LabelEncoder with OneHotEncoder for encoding categorical data. Discuss one advantage and one limitation of OneHotEncoder.

Explain feature scaling with respect to data preprocessing.

Write and explain python code snippet to create and print a dataframe for Employee_Salary dataset.

Write a python script which creates and prints a series for Employee_Experience.

Assume that you have an existing dataset of data related to patient symptoms and your task is to make a prediction to diagonize if a patient is infected from the disease lung cancer or not. There are seven symptoms in the dataset which have only two values, 1 for yes and 0 for no. and the diagnosis field has the values infected and not_infected only. Write a python script create a pre-processing template which includes finding missing data, categorizing and encoding the field gender.

Dataset: Patient_Record (ID, pateintname, gender, age, severe_cough, chest_pain, weight_loss, weakness, reoccuring_bronchitis, bonepain, seizure, Diagnosis)

Example: (P001, Arjun Dave, M, 43, 1,0,1,1,1,0,1,Infected)

Unit-2

=====

Draw the diagram and explain simple linear regression.

What is regression? Explain best fitting line in the context of simple linear regression.

Explain multiple linear regression with appropriate example.

Draw flow chart and explain backward elimination method for multiple linear regression.

What is multiple linearity and dummy variable trap? Explain backward elimination methods with respect to multiple linear regression.

In Y=bo + b1x, what is the importance of bo and b1? What is best fitting line? Discuss using graph.

List out steps for bidirectional elimination method for building a multiple linear regression model. If we will consider all possible multiple linear regression models, how many models are possible for dataset having 10 columns?

Explain backward elimination and forward selection method for building a multiple linear regression model.

Below are the 10 records for column state. In multiple linear regression, how many dummy variables one need to add in this example? Explain.

(State: Gujarat, Maharashtra, Bihar, Goa, Gujarat, Goa, Maharashtra, Gujarat, Goa, Bihar)

If p value of X variable is 0.06 and your significance level is 5%, will you keep X variable in model? Why?

Why one need to remove some of the independent variables while creating model? Write down five methods for building a model in multiple linear regression and explain any one in detail.

How can we add all dummy variables in multiple linear regression equations? Write usage of it.

Write the situation where polynomial regression is needed. Explain polynomial regression with proper example.

What is linear regression? When one need to use polynomial regression? Why it is called linear?

Discuss Decision Tree and Random Forest regression model.

What is CART? In which situations, decision tree regression model is best suited and why?

Random forest uses boosting or bagging technique? Explain boosting and bagging with respect to ensemble learning.

Draw flow chart and explain random forest regression algorithm.

Prepare a regression template using python script which can be used for Decision Tree.

Draw the diagram and explain the creation of decision tree regression model.

Assume that you have an existing dataset of data for one month related to posts on social network. Classify the friends of a person into friends having positive, negative and neutral thinking using decision tree regression and also plot the data using graph.

Dataset: User_Post (FriendID, Name, LoPP , LoNP, LoNuP, DLoPP,

DLoNP, DLoNuP, FoPC,FoNC, FoNuC, FoPPS, FoNPS,Prediction)

Example: (F001,Vikram Mehta, 110,380,40, 410, 180,

100,250,300,200,480,Negative)

LoPP= likes_on_positive_post, LoNP=likes_on_negative_post, LoNuP = likes_on_neutral_posts, DLoPP= dislike_on_positive_post, DLoNP = dislike_on_negative{_post, DLoNuP = dislike_on_neutral_post, FoPC= freqeuncy_of_positive_comments,FoNC = frequency_of_negative_comments, , FoNuC= frequency_of_neutral_comments, FoPPS= frequecny_of_positive_post_sharing, FoNPS = frequency_of_negative_post_sharing.

Unit-3

=====

Write and explain python code snippet for logistic regression.

Write and explain equations of sigmoid function and logistic regression.

Develop a classification template using python code for K-NN algorithm.

What is Euclidian distance? Write down and discuss steps of KNN algorithm.

Write python code snippet and explain Support Vector Machine as a classifier.

Explain Support vector machine algorithm with a proper graphical representation and mathematical notations.

What is a support vector? How SVM is different from other classifiers?

How support vector machine works? Discuss.

Write and explain python code snippet for Naive Bayes classifier.

Discuss Bayes theorem with example.

When to use naive Bayes classifier? Discuss with appropriate examples.

What is the role of conditional probability in Naive Bayes? Discuss using proper example.

Assume that you have the following data set. Calculate and find the individual conditional probabilities for each feature and calculate the final probability of whether a match shall be scheduled to play or not using Naïve Bayes.

Outlook	Temperature
Sunny	Hot
Rainy	Mild
Overcast	Cold
Rainy	Cold
Sunny	Hot
Overcast	Mild
Rainy	Mild

Write the steps to create a machine learning model and explain each step in detail.

Ram is having 600 Kesar and 400 Rajapuri Mango. Out of all, 500 ripe and 500 unripe. Out of all unripe mango, Kesar mango are 45%. Find the probability that the selected Kesar mango is Ripe using Bayes' theorem.

In Eru village, 40 Neem trees and 60 Peepal trees are available. Out of 100 trees, 40% leaves are green and other are brown. Out of all brown leaves, 60% are from Peepal trees. Calculate the probability that the selected leaf is green and of Neem tree.

Write short note on K-NN classifier.

Write short note on SVM classifier.

Write a detailed note on KNN classifier. Also discuss how to calculate the value of K.

Write short note on logistic regression classifier.

Why the term naive is used in Bayes classification method? What is P(X) and what if more than two features are available in naive Bayes classification method?

What do you mean by prior probability, marginal likelihood, likelihood and posterior probability. Show the calculation for all of them using one example.

Differentiate K-NN and SVM.

Differentiate between logistic regression classifier and support vector machine.

Unit-4

=====

Explain working of clustering and association as unsupervised machine learning.

List four clustering categories and explain any two.

Explain density based and hierarchical based clustering.

Compare partitioning based and grid based clustering. Explain both methods.

Draw the flow chart and explain K-Means clustering.

Draw diagrams and explain how K-Means clustering algorithm is sensitive to outliers and initial seeds.

Discuss any two methods to choose value of K in K-Means clustering algorithm.

Write python code snippet and explain elbow method in K-Means clustering algorithm. Draw suitable diagram.

Draw the flow chart and step-by-step explain working of agglomerative hierarchical clustering.

What is dendrogram? Explain how to select number of clusters in hierarchical clustering using dendrogram.

Explain support and confidence metrices with respect to measure association.

Step-by-step explain Apriori algorithm.

With diagram explain, minimum linkage clustering and centroid linkage clustering.

Explain maximum linkage clustering and average linkage clustering with suitable diagram.

Unit-5

=====

What do you mean by reinforcement learning? How does it differ from other approximation methods?

Draw the diagram and explain reinforcement learning.

With example, explain exploration versus exploitation with respect to reinforcement learning.

With appropriate example, explain multi-armed bandit problem.

Explain working of upper confidence bound algorithm with appropriate diagram and equation.

Write and explain python code snippet to implement upper confidence bound.

How to solve multi-armed bandit problem with Thompson sampling? Discuss.

Explain Thompson sampling algorithm with appropriate python code snippet.

What is a Neural Network (NN)? Which types of problems are suitable to solve using NN? Discuss.

Explain Hidden Layer with suitable example.

With a suitable example explain back propagation in Neural Network.

Explain artificial neural network with the help of a proper diagram.

Why deep learning was not appreciated initially? Describe read-write speed, data retention, power usage and data density for different storage media. Discuss processing capacity in the context of time line.

Which activation functions are commonly applied in hidden layer and output layer? Discuss them in detail.

Write and discuss python script for fitting the CNN to the images.

What do you mean by compiling an CNN? Write and discuss python script for the same. What is image augmentation? Why one need to use it?

What is pooling and why one need it? Discuss pooling with the help of example feature map.

What do you mean by back propagation? What are the advantages of it? List out steps for training ANN.

What do you mean by compiling an ANN? Write and discuss python script for the same. Write and discuss python script for predicting test results and making the confusion matrix in context of ANN.

What is activation function? List out any four of them and explain any two in detail. If dependent variables having value 0 or 1, which activation function is more suitable? Why?

Draw the basic architecture of CNN and explain it in detail.

Explain convolution operation in detail with one 7x7 binary image and 3x3 kernel/feature detector. How feature map is useful for image understanding? Also discuss any four filters/kernels/feature detectors.

What do you mean by curse of dimensionality? In which situation stochastic gradient descent needed? Write two basic differences between normal gradient descent and stochastic gradient descent.

What is neurons? Draw the structure of single neuron. Discuss in detail.

Differentiate standardization and normalization. What is weight in the context of neural network?

Who is Geoffrey Hinton? Discuss the popularity reasons of deep learning now-a-days. How do the neural networks work? Discuss with example.

Write python script for making ANN.

Differentiate gray/black and white images with color images. List out steps of convolutional NN and discuss ReLU layer of CNN in detail.

Write short note on ANN.

Discuss activation functions in detail.

Draw basic architecture of CNN and explain any two phase in detail.

List different types of pooling methods. What is the importance of pooling layer? How it works?

Discuss the process of image classification using CNN.

Which are the different activation functions? With one example, discuss the process of NN learning.

Unit-6

=====

What is dimensionality reduction? Why it is needed? List out the methods for doing so.

Discuss dimensionality reduction in detail with suitable example.

Write python script for implementing PCA and discuss.

What makes PCA, an unsupervised model? What are the advantages of PCA? At which position one needs to apply PCA for classification problem?

In which situation Linear Discriminant Analysis is best suited? Discuss with suitable example.

Write python script for implementing Linear Discriminant Analysis and discuss it.

Discuss advantage and limitations of LDA compare to kernel PCA.

Compare advantages of PCA versus LDA.

What is kernel? Discuss different types of kernel in the context of machine learning. How Kernel PCA is better than PCA?

How Kernel PCA is different from PCA? Discuss.

Explain working of Kernel PCA.

Write and explain python code snippet for kernel PCA.

List out the limitations of Principal component analysis. How kernel PCA is different from PCA?

Write short note on PCA.

Discuss LDA in detail.

List out any two dimensionality reduction technique and explain any one in detail.

Compare LDA and Kernel PCA.

What makes LDA, a supervised model? Differentiate PCA and LDA.

Differentiate PCA and Kernel PCA.

Why one need dimensionality reduction? List out linear and non linear methods. Which one is better in which situation and why?

Which one is better dimensionality reduction technique and why? Write python script of it.

Write python script for implementing Kernel PCA and discuss it.

Compare PCA, LDA and Kernel PCA.

Which feature extraction technique for dimensionality reduction, works on non linear data? Explain it in detail.

Compare and explain python script for PCA and Kernel PCA.

Write short note on LDA.

LDA is better compared to kernel PCA. True/False? Justify.

What is the importance of dimensionality reduction? Write down the snippet code of Linear Discriminant Analysis.