

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Ramapuram Campus, Bharathi Salai, Ramapuram, Chennai - 600089

FACULTY OF ENGINEERING AND TECHNOLOGY

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING



QUESTIONBANK

DEGREE / BRANCH: B.TECH CSE WITH SPECIALIZATION AIML

VI SEMESTER

18CSE479T – STATISTICAL MACHINE LEARNING

Regulation–2018

AcademicYear2021-22

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

Ramapuram Campus, Bharathi Salai, Ramapuram, Chennai-600089

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

SUBJECT :18CSE479T – Statistical Machine Learning SEM/YEAR : VI / III			
Course Outcomes: CO1: Acquire the knowledge on statistical machine learning techniques. CO2: Acquire the ability to build model based on logistic regression and random forest techniques CO3: Understand the basic ideas of probability and work on probabilistic approaches like Naïve Bayes, Bayes Theorem CO4: Apply the knowledge of Kernel functions in practical applications CO5: Apply the knowledge of K-means clustering on real world examples CO6: Acquire the knowledge on using PCA and SVD with Scikit-learn			
UNIT I			
Statistical terminology for model building and validation-Machine Learning, Major differences between statistical modeling and machine learning- Steps in machine learning model development and deployment - Statistical fundamentals and terminology for model building and validation - Bias versus variance trade-off,Train and test data - Linear regression versus gradient descent Machine learning losses- When to stop tuning machine learning models - When to stop tuning machine learning models - Train, validation, and test data Cross-validation – Grid Search – Machine Learning model overview			
PART-A (Multiple Choice Questions)			
Q.No	Questions	Course Outcome	Competence BT Level
1	_____ is widely used and effective machine learning algorithm based on the idea of bagging. a) Regression b) Classification c) Decision Tree d) Random Forest Answer: d	1	1
2	Define Machine Learning a) It is the branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of numerical data. b) Applying statistics on data to find underlying hidden relationships by analyzing the significance of the variables. c) It is the branch of computer science that utilizes past experience to learn from and use its knowledge to make future decisions. d) It is the collection of entire data. Answer: c	1	1
3	Categories of unsupervised learning a) Regression problem b) Clustering c) Classification d) Monte Carlo Method Answer: b	1	2
4	Reinforcement Learning is closer to the a) Machine Learning b) Deep Learning c) Artificial Intelligence d) Supervised learning Answer: c	1	2
5	Which of the following is not a supervised learning? a) Naïve Bayesian b) PCA	1	1

	c) Linear Regression d) Decision Tree Answer Answer: b		
6	Real-Time decisions, Game AI, Learning Tasks, Skill Aquisition, and Robot Navigation are applications of which of the following a) Supervised Learning: Classification b) Reinforcement Learning c) Unsupervised Learning: Clustering d) Unsupervised Learning: Regression Answer: b	1	3
7	What is the most significance phase in a genetic algorithm? a) Selection b) Mutation c) Crossover d) Fitness Function Answer: c	1	1
8	What is Machine Learning? (i) Artificial Intelligence (ii) Deep Learning (iii) Data Statistics a) Only (i) b) (i) and (ii) c) All d) None Answer: b	1	1
9	Targetted marketing, Recommended Systems, and Customer Segmentation are applications in which of the following a) Supervised Learning: Classification b) Unsupervised Learning: Clustering c) Unsupervised Learning: Regression d) Reinforcement Learning Answer: b	1	3
10	What strategies can help reduce overfitting in decision trees? (i) Enforce a maximum depth for the tree (ii) Enforce a minimum number of samples in leaf nodes (iii) Pruning (iv) Make sure each leaf node is one pure class a) All b) (i), (ii) and (iii) c) (i), (iii), (iv) d) None Answer: b	1	1
11	Data will be split into 50:25:25 illustrate it a) 50% testing; 25% training; 25% validation b) 50% training; 50% testing c) 50% training; 25% validation; 25% testing d) 50% validation; 25% training; 25% testing Answer: c	1	3
12	Define population a) A small portion of observation b) The various bits and pieces involved in the observation c) One or more variables involved in the observation d) The complete list of observations Answer: d	1	1
13	>>> train_data,test_data = train_test_split(original_data,train_size = 0.7, random_state=42) From the above code, What is the percentage of validation and testing of the data? a) 30%	1	3

	b) 70% c) 42% d) 60% Answer: a		
14	Identify the model is having high bias a) Over fitting b) Worst fitting c) Best fitting d) Under fitting Answer: d	1	2
15	Define the Descriptive statistics used to summarize the data a) Mean, standard deviation for continuous data sets b) Hypothesis testing for continuous data sets c) Inferences drawn over the discrete data sets d) Correlation relationships between the data Answer: a	1	1
16	Which of the following is a widely used and effective machine learning algorithm based on the idea of bagging? a) Decision Tree b) Random Forest c) Regression d) Classification Answer: b	1	1
17	Which of the following is a disadvantage of decision trees? a) Decision trees are prone to be overfit b) Decision trees are robust to outliers c) Factor analysis d) None of the above Answer: a	1	1
18	If Linear regression model perfectly first i.e., train error is zero, Analyze it a) Test error is also always zero b) Test error is non zero c) Couldn't comment on Test error d) Test error is equal to Train error Answer: c	1	3
19	Which of the following metrics can be used for evaluating regression models? i) R Squared ii) Adjusted R Squared iii) F Statistics iv) RMSE / MSE / MAE a) ii and iv b) i and ii c) ii, iii and iv d) i, ii, iii and iv Answer: d	1	1
20	How many coefficients do you need to estimate in a simple linear regression model (One independent variable)? a) 1 b) 2 c) 3 d) 4 Answer: b	1	1
21	Which of the following statement(s) is / are true for Gradient Decent (GD) and Stochastic Gradient Decent (SGD)? 1. In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function. 2. In SGD, you have to run through all the samples in your training set for a single update of a parameter in each iteration.	1	1

	<p>3. In GD, you either use the entire data or a subset of training data to update a parameter in each iteration.</p> <p>a) Only 1 b) Only 2 c) Only 3 d) 1 and 2 e) 2 and 3 f) 1,2 and 3</p> <p>Answer: a</p>		
22	<p>Which of the following hyper parameter(s), when increased may cause random forest to over fit the data?</p> <p>1. Number of Trees 2. Depth of Tree 3. Learning Rate</p> <p>a) Only 1 b) Only 2 c) Only 3 d) 1 and 2 e) 2 and 3 f) 1,2 and 3</p> <p>Answer: b</p>	1	1
23	<p>Analyze the wrong combination</p> <p>a) True negative = correctly rejected b) False negative = correctly rejected c) False positive = correctly identified d) all of the above</p> <p>Answer: c</p>	1	3
24	<p>Which of the following is the categorical outcome</p> <p>a) RMSE b) RSquared c) Accuracy d) all of the above</p> <p>Answer: c</p>	1	1
25	<p>For k cross validation, smaller k value implies more variance. Analyze it.</p> <p>a) True b) False</p> <p>Answer: a</p>	1	3
PART B (4 Marks)			
1	Analyze the path from statistical to machine learning	1	3
2	Explain Confusion matrix	1	2
3	Illustrate Supervised learning with real time examples	1	2
4	Classify the machine learning losses	1	2
5	When to stop tuning machine learning models	1	3
6	Explain Cross Validation	1	2
7	Illustrate test, validation and test data	1	2
8	Explain Grid search	1	2
PART C (12 Marks)			
1	Compare Statistical and Machine learning	1	2
2	Illustrate the Machine learning terminology for model building and validation	1	2

3	Explain in detail category of machine learning	1	2
4	Explain Quantiles and it types in detail	1	2
5	Compare Linear regression with gradient descent	1	3

UNIT II			
Comparison between regression and machine learning models- Compensating factors in machine learning models- Assumptions of linear regression Steps applied in linear regression modelling - Machine learning models - ridge and lasso regression - Example of ridge regression machine learning, Example of lasso regression machine learning mode - Logistic Regression Versus Random Forest-Maximum likelihood estimation - Terminology involved in logistic regression - Applying steps in logistic regression modelling - Random forest - Example of random forest using German credit data - Grid search on random forest - Variable importance plot - Comparison of logistic regression with random forest			
PART-A (Multiple Choice Questions)			
Q.No.	Questions	Course Outcome	Competence BT Level
1	Which of the following methods do we use to find the best fit line for data in Linear Regression? A) Least Square Error B) Maximum Likelihood C) Logarithmic Loss D) Area Under Curve	2	1
2	Which of the following evaluation metrics can be used to evaluate a model while modeling a continuous output variable? A) AUC-ROC B) Accuracy C) Logloss D) Mean-Squared-Error	2	1
3	Suppose that we have N independent variables ($X_1, X_2 \dots X_n$) and dependent variable is Y. Now Imagine that you are applying linear regression by fitting the best fit line using least square error on this data. You found that correlation coefficient for one of it's variable(Say X_1) with Y is -0.95. Which of the following is true for X_1 ? A) Relation between the X_1 and Y is weak B) Relation between the X_1 and Y is strong C) Relation between the X_1 and Y is neutral D) Correlation can't judge the relationship	2	3
4	We can also compute the coefficient of linear regression with the help of an analytical method called "Normal Equation". Which of the following is/are true about Normal Equation? 1. We don't have to choose the learning rate 2. It becomes slow when number of features is very large 3. There is no need to iterate A) 1 and 2 B) 1 and 3 C) 2 and 3 D) 1,2 and 3	2	3
5	What will happen when you apply very large penalty? A) Some of the coefficient will become absolute zero B) Some of the coefficient will approach zero but not absolute zero C) Linear regression is sensitive to outliers D) Linear regression is not sensitive to outliers	2	2
6	In terms of bias and variance. Which of the following is true when you fit degree 2 polynomial?	2	2

	A) Bias will be high, variance will be high B) Bias will be low, variance will be high C) Bias will be high, variance will be low D) Bias will be low, variance will be low		
7	Now we increase the training set size gradually. As the training set size increases, what do you expect will happen with the mean training error? A) Increase B) Decrease C) Remain constant D) Can't Say	2	3
8	What do you expect will happen with bias and variance as you increase the size of training data? A) Bias increases and Variance increases B) Bias decreases and Variance increases C) Bias decreases and Variance decreases D) Bias increases and Variance decreases	2	3
9	What would be the root mean square training error for this data if you run a Linear Regression model of the form ($Y = A_0 + A_1X$)? A) Less than 0 B) Greater than zero C) Equal to 0 D) Greater than or equal to zero	2	2
10	Suppose you got the tuned hyper parameters from the previous question. Now, Imagine you want to add a variable in variable space such that this added feature is important. Which of the following thing would you observe in such case? A) Training Error will decrease and Validation error will increase B) Training Error will increase and Validation error will increase C) Training Error will increase and Validation error will decrease D) Training Error will decrease and Validation error will decrease	2	1
11	Regression models a target prediction value based on _____. A) dependent variable B) independent variables C) independent value D) dependent value	2	1
12	Which of the following is a disadvantage of non-parametric machine learning algorithms? a) Capable of fitting a large number of functional forms (flexibility) b) Very fast to learn (speed) c) More of a risk to overfit the training data (overfitting) d) They do not require much training data	2	1
13	The scikit-learn Python machine learning library provides an implementation of the Ridge Regression algorithm via the Ridge class. Confusingly, the lambda term can be configured via the “_____” argument when defining the class. a) Lambda. b) Gamma. c) Beta. d) Alpha.	2	1
14	Ridge regression can reduce the slope close to zero (but not exactly zero) but Lasso regression can reduce the slope to be exactly equal to zero. a) Both statements are True about Ridge and Lasso. b) Both statements are False about Ridge and Lasso. c) True statement about Ridge but not about Lasso. d) True statement about Lasso but not about Ridge.	2	2

15	<p>The following statement is</p> <p>I. Lasso Regression helps to reduce overfitting and it is particularly useful for feature selection.</p> <p>II. Lasso regression can be useful if we have several independent variables that are useless.</p> <p>a) Statement (I) is true and statement (II) is false. b) Statement (I) is false and statement (II) is true. c) Both Statement (I) & (II) are wrong. D) Both Statement (I) & (II) are true.</p>	2	1
16	<p>Which of the following is used where the target variable is of categorical nature?</p> <p>A. Keras B. Knime C. Logistic Regression D. MXNet</p>	2	1
17	<p>How many different types of Logistic Regression?</p> <p>A. 2 B. 3 C. 4 D. 5</p>	2	1
18	<p>_____ are defined as the ratio of the probability of an event occurring to the probability of the event not occurring.</p> <p>A. Simple B. Even C. Regex D. Odds</p>	2	1
19	<p>Logistic regression is used when you want to:</p> <p>A. Predict a dichotomous variable from continuous or dichotomous variables. B. Predict a continuous variable from dichotomous variables. C. Predict any categorical variable from several other categorical variables. D. Predict a continuous variable from dichotomous or continuous variables.</p>	2	2
20	<p>_____ helps improve machine learning results by combining several models.</p> <p>A) Machine Learning B) bagging C) Entropy D) Ensemble learning</p>	2	1
21	<p>Which of the following is statistical boosting based on additive logistic regression?</p> <p>a) gamBoost b) gbm c) ada d) mboost</p>	2	1
22	<p>The measures developed for selecting the best split are often based on the degree of impurity of the child nodes. Which of the following is NOT an impurity measure?</p> <p>a) Gini b) Entropy c) Pruning d) Classification error</p>	2	1
23	<p>We are dealing with samples x where x is a single value. We would like to test two alternative regression models:</p> <p>1) $y = ax + e$</p>	2	3

	2) $y = ax + bx^2 + e$ Which of the two models is more likely to fit the test data better? a) model 1 b) model 2 c) both will equally fit d) impossible to decide		
24	In the general case, imagine that we have d binary features, and we want to count the number of features with value 1. How many leaf nodes would a decision tree need to represent this function? a) 2^1 leaf nodes b) 2^d leaf nodes c) 2^{d-1} leaf nodes d) $2^d - 1$ leaf nodes	2	2
25	Which of the following classifiers can generate linear decision boundary? a) Linear SVM b) Random forest c) Naïve Bayes d) k-NN	2	1
PART B (4 Marks)			
1	Compensating factors in machine learning models	2	2
2	Explain the steps applied in linear regression modeling	2	2
3	Compare Ridge and Lasso regression modeling	2	2
4	Explain Logistics Regression and its advantages	2	2
5	Explain the steps applied in logistic regression modeling	2	2
6	Write short notes on Random Forest	2	2
7	Explain random forest using German credit data	2	2
8	Explain grid search on random forest	2	2
PART C (12 Marks)			
1	Compare Linear regression and Machine learning model	2	2
2	Assumption of linear regression in detail	2	3
3	Explain simple linear regression from first principles	2	2
4	Explain Maximum likelihood estimation	2	2
5	Terminology involved in logistic regression	2	1

UNIT III			
K-nearest neighbors-KNN voter example - Curse of dimensionality-Curse of dimensionality with 1D, 2D, and 3D example. Curse of dimensionality with 3D example. KNN classifier with breast cancer Wisconsin data example- Naive Bayes - Probability fundamentals - Joint probability- Understanding Bayes theorem with conditional probability- Naive Bayes classification- Laplace estimator- Naive Bayes SMS spam classification example.			
PART-A (Multiple Choice Questions)			
Q. No	Questions	Course Outcome	Competence BT Level
1	Time to classify a new example than with a model in KNN requires? a. Depends on Data b. More Time c. None of these d. Less time	3	2
2	The Euclidean distance between two a set of numerical attributes is called as? a. Closeness b. Validation data c. Error Rate d. testing data	3	1
3	Which capture the local structure in data ? a. Low Value b. High Value c. Error Rate d. Median value	3	1
4	Which provide more smoothing, less noise ? a. Low Value b. High Value c. Error Rate d. Median value	3	1
5	Which of the following option is true about k-NN algorithm? A) It can be used for classification B) It can be used for regression C) It can be used in both classification and regression	3	1
6	Which of the following machine learning algorithm can be used for imputing missing values of both categorical and continuous variables? A) K-NN B) Linear Regression C) Logistic Regression	3	1
7	Which of the following distance measure do we use in case of categorical variables in k-NN? 1.Hamming Distance 2. Euclidean Distance 3. Manhattan Distance A) 1 B) 2 C) 3 D) 1 and 2 E) 2 and 3 F) 1,2 and 3	3	1
8	What is the naïve assumption in a Naïve Bayes Classifier A. all the classes are independent of each other B. all the features of a class are independent of each other	3	2

	<p>C. the most probable feature for a class is the most important feature to be considered for classification</p> <p>D. all the features of a class are conditionally dependent on each other</p>	3	3
9	<p>Spam Classification is an example for ?</p> <p>a. Naive Bayes</p> <p>b. Probabilistic condition</p> <p>c. Random Forest</p>	3	2
10	<p>Time complexity for Naive Bayes classifier for n feature, L class data is</p> <p>a. $n * L$</p> <p>b. $O(n+L)$</p> <p>c. $O(n * L)$</p> <p>d. $O(n/L)$</p>	3	3
11	<p>A list of symptoms, predict whether a patient has disease X or not</p> <p>a. Medical Diagnosis</p> <p>b. Weather Diagnosis</p> <p>c. Spam Diagnosis</p>	3	2
12	<p>How many Firms in Exact Bayes Calculation ?</p> <p>a. One</p> <p>b. Two</p> <p>c. Three</p> <p>d. Four</p>	3	1
13	<p>In Naive Bayes the relation between $P(\text{Fraud})$ and $P(\text{Truthful})$ is ?</p> <p>Options :</p> <p>a. Greater than</p> <p>b. Lesser than</p> <p>c. Equal to</p> <p>d. Equal to 1</p>	3	3
14	<p>If a bag contains red and blue balls and randomly picked 10 balls one by one with replacement and out of 10, 3 red balls appear in trials, then the probability of red ball is</p> <p>a. 0.6</p> <p>b. 0.3</p> <p>c. 0.4</p> <p>d. 0.5</p>	3	3
15	<p>If a trial has two outcomes then the events are considered as _____.</p> <p>a. double events</p> <p>b. success event</p> <p>c. mutually exclusive</p> <p>d. failure event</p>	3	2
16	<p>Naïve Bayes is _____</p> <p>a. Conditional Independence</p> <p>b. Conditional Dependence</p> <p>c. Both a and b</p> <p>d. None of the above</p>	3	1
17	<p>Spam Classification is an example for</p> <p>a. Probabilistic condition</p> <p>b. Random forest</p> <p>c. Naive bayes</p> <p>d. All the above</p>	3	2
18	<p>Which preprocessing technique in NLP used to reduce the dimensionality?</p> <p>a. Word tokenization</p> <p>b. Lemmatization of words</p>	3	2

	c. Part-of-Speech tagging d. Removal of punctuation		
19	Naïve Bayes requires _____ a. Categorical values b. Numerical values c. Either a or b d. Both a and b	3	1
20	The probability of an event before evidence is seen is _____ a. Posteriori probability b. Prior probability c. Likelihood d. conditional probability	3	1
21	_____ is the likelihood of an outcome occurring based on the previous outcome occurring. a. Posteriori probability b. Prior probability c. marginal probability d. conditional probability	3	1
22	_____ is a statistical measure that evaluates how relevant a word is to a document in a collection of documents. a. Term frequency b. Inverse document frequency c. Term frequency-Inverse document frequency d. document frequency	3	1
23	Which function describes the joint probability of the observed data as a function of the parameters of the chosen statistical model? a. Likelihood b. Independence c. Dependence d. Marginal	3	1
24	Which package is used for all preprocessing stages in Natural Language Processing? a. stats b. nltk c. csv d. nlp	3	1
25	Which technique is used to tackle the problem of zero probability in Naïve Bayes machine learning algorithm? a. Simple exponential b. Random Walk c. Moving Average d. Laplace estimator	3	1
PART B (4 Marks)			
1	Explain short notes on KNN	3	2
2	Explain KNN voter example	3	2
3	Explain Curse of dimensionality	3	2
4	What is tuning of k-value in KNN classifier	3	2
5	What is a Joint probability? Explain it with a simple Venn Diagram.	3	1
6	Explain the use of Conditional probability.	3	1
7	What is the purpose of using Laplace estimator?	3	2

8	Explain TF-IDF with an example.	3	3
PART C (12 Marks)			
1	What will happen if the Curse of dimensionality with 1D, 2D, and 3D Explain	3	2
2	KNN classifier with breast cancer Wisconsin data example	3	3
3	Describe the various preprocessing stages for Natural Language Processing with a suitable example.	3	1
4	Explain Naïve Bayes Classification with an example	3	2
5	Discuss SMS Spam classification using Naïve Bayes algorithm.	3	3

UNIT IV			
Support Vector Machines and Neural Networks-Support vector machines working principles-Maximum margin classifier- Support vector classifier- Support vector machines- Kernel functions- Artificial neural networks – ANN - Forward propagation and back propagation-Optimization of neural networks-Stochastic gradient descent – SGD- Introduction to deep learning-Solving methodology- Deep learning software			
PART-A (Multiple Choice Questions)			
Q. No	Questions	Course Outcome	Competence BT Level
1	Closest Point to the hyper plane are_____ a. super vectors b. Linear vectors c. Non linear vectors d. None of these	4	1
2	SVM, which best segregates classes into how many classes? a. 1 b. 2 c. 3 d. 4	4	2
3	SVM is a supervised <u>Machine Learning</u> can be used for a. Regression b. Classification c. Either a or b d. None of the above	4	2
4	Which is used to maximize the margin between the two classes? a. minimum classifier b. maximum separator c. minimum separator d. maximum classifier	4	1
5	The observations touching both the positive and negative hyperplanes called as a. maximum classifier b. support vector c. minimum classifier d. minimum vector	4	1
6	Which can be used to work with non-linearly separable data? a. support vector classifier b. minimum classifier c. maximum classifier d. kernel trick	4	2

7	In which function, given the original feature vectors, return the same value as the dot product of its corresponding mapped feature vectors? a. Kernel functions b. SVM functions c. Hyperplane functions d. Linear functions	4	2
8	Which kernels are popularly used, especially with degree 2? a. Mononomial Kernel b. Polynomial Kernel c. Trinomial Kernel d. Multinomial Kernel	4	1
9	The hyperplane with the maximum margin of separation width is called as _____ a. Maximum margin classifier b. Minimum margin classifier c. Kernel function d. Super vectors	4	1
10	Which value of gamma in RBF Kernel will give you low bias and high variance solutions? a. 0 b. 1 c. high d. low	4	2
11	In RBF Kernel, which value of the gamma value gives you a pointed bump in the higher dimensions? a. larger b. smaller c. medium d. 0	4	2
12	Which function takes a real valued number and squashes into a range between 0 and 1? a. tanh b. Relu c. linear d. sigmoid	4	1
13	Which model the relationship between a set of input signals and output signals using a model derived from a replica of the biological brain, which responds to stimuli from its sensory inputs? a. Artificial neural networks (ANNs) b Artificial Intelligence c. Brain d. Human interface	4	1
14	Which activation function is used in Linear regression? a. tanh b. Relu c. linear d. sigmoid	4	1
15	The number of neurons in the input layer is based on _____ a. dependent variables b. independent variables c. weight d. all the above	4	2
16	Which deep learning software is Google's deeplearninglibrary running on top of Python/C++ a. Caffe	4	1

	b. Keras c. Theano d. TensorFlow		
17	In which technique, all training observations are utilized in each iteration for optimizing the weights of neural networks? a. Stochastic gradient descent b. Adadelta c. Adaptive moment estimation d. Batch gradient descent	4	2
18	Which utilizes neural networks for building models to solve both supervised and unsupervised problems on structured and unstructured datasets? a. Deep Learning b. ANN c. SVM d. All the above	4	2
19	Which method is used to solve deep layers by calculating the error of the network at output units and propagate back through layers to update the weights to reduce error terms? a. Forward propagation b. Backward propagation c. Bi-propagation d. Tri-propagation	4	1
20	In the thumb rule in designing Deep Neural networks, all hidden layers should have the _____ number of neurons per layer. a. 1 b. 2 3. same 4. different	4	1
21	The number of neurons in the output layer is decided by the number of _____ the model needs to be predicted. a. independent variables b. weight c. classes d. dependent variables	4	1
22	Which is a Python-based deep learning library developed by the University of Montreal? a. Theano b. Caffe c. Keras d. TensorFlow	4	1
23	Which thumb rule improves convergence, in addition to the use of momentum and dropout in designing deep neural networks? a. two hidden layers are good enough to solve the majority of problems b. Reduction in step size after each iteration c. All hidden layers should have the same number of neurons per layer d. Reduction in step size after each iteration in each layer	4	2
24	Which deep learning software is a deep learning library primarily used for processing pictures? a. Tensorflow b. Caffe c. Keras d. Lasagne	4	1
25	In _____ technique, one observation per iteration is considered for optimizing the weights of neural networks? a. Stochastic Gradient Descent	4	2

	b. Adadelta c. Adaptivemomentestimation d. Batch gradient descent		
PART B (4 Marks)			
1	When the Support Vector Machine are used? Give suitable example to illustrate it.	4	3
2	Write short notes on different types of Kernel function.	4	1
3	Explain Super Vector Classifiers with an example.	4	3
4	What are the parameters requiredforchoosingforbuildingneuralnetworks?	4	2
5	What is an Activation function and what are its types?	4	1
6	Write the Thumb rules in designing deep neural networks.	4	1
7	Give the various types of deep learning software used by manypractitionersacross theworld.	4	2
8	Explain the working principle of Artificial Neural Network.	4	2
PART C (12 Marks)			
1	Discuss on Super Vector Machine working principles in detail.	4	1
2	Describe in detail on Kernel functions.	4	1
3	Explain in detail on Forward and Backward Propagation in Artificial Neural Network with a neat diagram.	4	2
4	Illustrate Stochastic gradient descent for Optimization of neural networks with a neat diagram.	4	2
5	Explain the Deep architecture of Neural networks. Give the thumb rules for designing the Deep neural Networks. Briefly explain about Deep Learning software	4	2

UNIT V			
K-means clustering-K-means working methodology from first principles- Optimal number of clusters and cluster evaluation - The elbow method- K-means clustering with the iris data example- Principal component analysis - PCA-PCA working methodology from first principles- PCA applied on handwritten digits using scikit-learn- Singular value decomposition – SVD- SVD applied on handwritten digits using scikit-learn			
PART-A (Multiple Choice Questions)			
Q. No	Questions	Course Outcome	Competence BT Level
1	Which of the following is required by K-means clustering? a. defined distance metric b. number of clusters c. initial guess as to cluster centroids d. all of the above	5	2
2	_____is the task of grouping observations in such a way that members of the samecluster are more similar to each other and members of different clusters are very differentfrom each other. a. Classification b. Regression c. Clustering d. PCA	5	1
3	Which method is used to determine the optimal number of clusters in k-meansclustering?	5	1

	a. Elbow b. Euclidean distance c. Manhattan distance d. Centroid		
4	What is the name of the measure of the compactness and separation of the clusters? a. k-means coefficient b. silhouette coefficient c. k-medoid coefficient d. k-centroid coefficient	5	1
5	Silhouette coefficient values range from a. -1 to +1 b. 0 to 1 c. -1 to 0 d. -1 to +2	6	1
6	Which reduces the dimensions of a dataset by projecting the data onto a lower-dimensional subspace? a. PCA b. Regression c. Clustering d. Classification	6	2
7	A 2D dataset could be reduced by projecting the points onto a _____ a. point b. plane c. cube d. line	6	1
8	What are called as the axes (directions) along which a linear transformation acts simply by stretching/compressing and/or flipping? a. Eigen values b. Eigen points c. Eigen vectors d. Eigen array	6	1
9	What is the importance of using PCA before the clustering? a. Find good features to improve clustering b. Avoid bad features c. Find the explained variance d. Find which dimension of data maximizes the features variance	6	2
10	In PCA, it transforms the variables into a new set of variables called as _____. a. linear components b. scalar components c. vector components d. principal components	6	1
11	Why is so important to standardize the data in PCA? a. Find the features which can predict Y b. Make the training time more fast c. Other people understand better d. Use the best practices of data wrangling	6	2
12	A three dimensional dataset could be reduced by projecting the points onto a _____ a. point b. plane c. cube d. line	6	1
13	Which is higher for compact clusters that are well separated and lower for	6	2

	overlapping clusters? a. silhouettecoefficient b. k-meanscoefficient c. k-medoid coefficient d. k-centroid coefficient		
14	Why do you have to drop unimportant features in PCA? a. find the correct clusters b. standardize the data c. To train the models faster d. Using the most important features will give better efficiency predicting the target.	6	2
15	K-means algorithm _____. a. Requires the dimension of the feature space to be no bigger than the number of samples b. has the smaller value of the objective function when k=1 c. minimizes the within class variance for a given number of clusters d. converges to the global optimum if and only if the initial means are chosen as some of the samples themselves	5	2
16	What does K refers in the K-means algorithm? a. fixed value b. complexity c. number of clusters d. number of iterations	5	2
17	Which of the following cases will K-means clustering give poor results? a. Data points with outliers b. Data points with different densities c. Data points with round shapes b. Data points with convex shapes	5	2
18	The equivalent of eigenvalues obtained throughthe SVD method are called as _____. a. singular values b. singular points c. singular vectors d. singular angles	6	1
19	Which is a measure of how much two variables change together and it is a measure ofthe strength of the correlation between two sets of variables? a. Precision b. recall c. Covariance d. SVD	6	1
20	When can you conclude that there will not be any correlation between two sets ofthevariables? a. Covariance = 0 b. Covariance =1 c. Covariance =2 d. Covariance =-1	6	2
21	SVD can be applied even on which type of matrices? a. Squared b. rectangular c. 1 dimensional d. 3 dimensional	6	1
22	Orthogonal matrices in SVD have _____ dimensions. a. same b. 1 c. 3	6	1

	d. different		
23	What is the name of the vector obtained equivalent to eigenvectors in SVD? a. singular values b. singular points c. singular vectors d. singular angles	6	1
24	The cost function of k-means is determined by which method? a. Elbow b. Euclidean distance c. Manhattan distance d. Centroid	5	1
25	The eigenvalues are defined only for _____ matrices. a. Squared b. rectangular c. 3 dimensional d. 1 dimensional	6	1
PART B (4 Marks)			
1	Discuss on evaluating the cost function of K-means clustering.	5	2
2	Explain the elbow method to determine the optimal number of clusters in k-means clustering.	5	1
3	Give few applications of K-means clustering.	5	2
4	Write notes on utilities of Principal Component Analysis.	6	2
5	Illustrate PCA for two Dimensional data with two principal components.	6	2
6	Illustrate the Graphical representation of Eigen values and Eigen vectors.	6	2
7	Write the advantages of Singular Value Decomposition	6	1
8	Explain the computation of singular values and singular vectors in SVD.	6	2
PART C (12 Marks)			
1	Explain K-means clustering working methodology from first principles with a suitable example.	5	2
2	Discuss in detail on K-means clustering with iris data example.	5	3
3	Describe Principal Component Analysis working methodology from first principles with a suitable example.	6	2
4	Illustrate the application of PCA for handwritten digits using scikit-learn.	6	3
5	Describe SVD applied on handwritten digits using scikit-learn.	6	3

Note:

1. **BT Level** – Blooms Taxonomy Level

2. **CO – Course Outcomes**

BT1 – Remember BT2 – Understand BT3 – Apply BT4 – Analyze BT5 – Evaluate BT6 – Create