Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

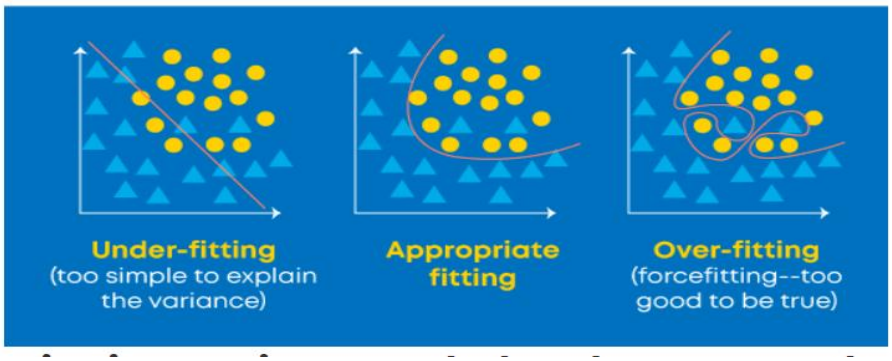
Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

Q. No.	
Q.1 a)	**Each question of 5 Marks**

i)Distinguish between supervised learning and Reinforcement learning. Illustrate with an example.

Distinguish w.r.t.
Definition, Type of problem, type of data, training,approach algorithm

ii) Why regularization come into play in Machine Learning?
At times when the model begins to underfit or overfit, regularization becomes necessary. It is a regression that diverts or regularizes the coefficient estimates towards zero. It reduces flexibility and discourages learning in a model to avoid the risk of overfitting. The model complexity is reduced and it becomes better at predicting.



iii)Why do RNNs struggle with long distance dependencies?
Because of vanishing gradients, the RNN doesn't learn the long-range dependencies across time steps. That means that there is a possibility that the word "what" and "time" are not considered when trying to predict the user's intention.
ii) Machine learning: Face recognition, handwritten recognition, credit card approval.
Not machine learning: calculate payroll, execute a query to database, use WORD

iv) Give three computer applications for which machine learning approaches seem appropriate and three for which they seem inappropriate.
Face recognition, handwritten recognition, credit card approval. Not machine learning: calculate payroll, execute a query to database, use WORD.

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

Q2a) Why is Linear Discriminant Analysis important? How do LDAs work? (5+5)

Linear discriminant analysis is primarily used here **to reduce the number of features to a more manageable number before classification**. Each of the new dimensions is a linear combination of pixel values, which form a template.

LDA works in a similar way to PCA. The aim of an LDA algorithm is to try to find the best linear combination that gives the maximum separation between the number of groups present. It calculates the discriminant scores from a linear combination of weights and centred data points.These weights are extracted from eigenvectors. Unlike PCA, the eigenvectors are not calculated from a covariance matrix but from a matrix which is computed from the transpose of the distance between groups multiplied by the distance between groups.

LDA can be performed in 5 steps:

1. Compute the mean vectors for the different classes from the dataset
2. Compute the scatter matrices (in-between-class and within-class scatter matrices).
3. Compute the eigenvectors and corresponding eigenvalues for the scatter matrices.
4. Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues.
5. Use this eigenvector matrix to transform the samples onto the new subspace.

Computing the mean vectors

First, calculate the mean vectors for all classes inside the dataset.

$$m_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i$$

Computing the scatter matrices

After calculating the mean vectors, the within-class and between-class scatter matrices can be calculated.

Within-class scatter matrix S_W

$$S_W = \sum_{i=1}^{c} S_i$$

where S_i is the scatter matrix for a specific class

$$S_i = \sum_{x \in D_i}^{n} (x - m_i)\ (x - m_i)^T$$

and m_i is the mean vector for that class

$$m_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i$$

Alternativeley the class-covariance matrices can be used by adding the scaling factor $\frac{1}{N-1}$ to the within-class scatter matrix.

$$\Sigma_i = \frac{1}{N_i - 1} \sum_{x \in D_i}^{n} (x - m_i)\ (x - m_i)^T$$

$$S_W = \sum_{i=1}^{c} (N_i - 1) \Sigma_i$$

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

$$m_1 = \frac{\sum_t w^T x^t r^t}{\sum_t r^t} = w^T m_1$$
$$m_2 = \frac{\sum_t w^T x^t (1 - r^t)}{\sum_t (1 - r^t)} = w^T m_2$$
$$s_1^2 = \sum_t (w^T x^t - m_1)^2 r^t$$
$$s_2^2 = \sum_t (w^T x^t - m_2)^2 (1 - r^t)$$

Mean and scatter after projection

$$J(\mathbf{w}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$ Good projection

b) i)Compare between Regression, classification and clustering

Regression	Classification	Clustering
• Supervised Learning	• Supervised Learning	• Unsupervised Learning
• Output is a continuous quantity	• Output is a categorical quantity	• Assigns data points into clusters
• Main aim is to forecast or predict	• Main aim is to compute the category of the data	• Main aim is to group similar items clusters
• Eg: Predict stock market price	• Eg: Classify emails as spam or non-spam	• Eg: Find all transactions which are fraudulent in nature
• Algorithm: Linear Regression	• Algorithm: Logistic Regression	• Algorithm: K-means

For the transaction shown in the table compute the following

i)Entropy of the collection of transaction records of the table w.r.t. classification

ii)What are the information gain of A1 ,A2 relative to the transaction of the table?

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

Instance	1	2	3	4	5	6	7	8	9
A1	T	T	T	F	F	F	F	T	F
A2	T	T	F	F	T	T	F	F	T
Target class	+	+	-	+	-	-	-	+	-

Formula of entropy , gain ,detail calculation with correct final answer(5m)

Formula, detail calculation and final answer of all following

Entropy of a1=09910

Gain(s,a1)=0.2295

Entropy of a2= 0.9838

Gain (s,a2)=0.0072

OR

ii)Can Random Forest Algorithm be used both for Continuous and Categorical Target Variables? What do you mean by Bagging?(2+8)

Yes, it can be used for both continuous and categorical target (dependent) variable. In random forest/decision tree, classification model refers to factor/categorical dependent variable and regression model refers to numeric or continuous dependent variable.

Bagging

Bootstrap Aggregation, also known as **Bagging**, is a powerful ensemble method that was proposed by Leo Breiman in 1994 to prevent overfitting. The concept behind bagging is to combine the

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

predictions of several base learners to create a more accurate output. Bagging is the application of the Bootstrap procedure to a high-variance machine learning algorithm, typically decision trees.

1. Suppose there are N observations and M features. A sample from observation is selected randomly with replacement (Bootstrapping).
2. A subset of features are selected to create a model with sample of observations and subset of features.
3. Feature from the subset is selected which gives the best split on the training data.
4. This is repeated to create many models and every model is trained in parallel
5. Prediction is given based on the aggregation of predictions from all the models.

This approach can be used with machine learning algorithms that have a high variance, such as decision trees. A separate model is trained on each bootstrap sample of data and the average output of those models used to make predictions. This technique is called bootstrap aggregation or bagging for short. Variance means that an algorithm's performance is sensitive to the training data, with high variance suggesting that the more the training data is changed, the more the performance of the algorithm will vary.

The performance of high variance machine learning algorithms like unpruned decision trees can be improved by training many trees and taking the average of their predictions. Results are often better than a single decision tree.

What Bagging does is help reduce variance from models that are might be very accurate, but only on the data they were trained on. This is also known as overfitting.

Overfitting is when a function fits the data too well. Typically this is because the actual equation is much too complicated to take into account each data point and outlier.

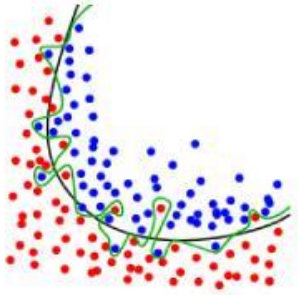
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

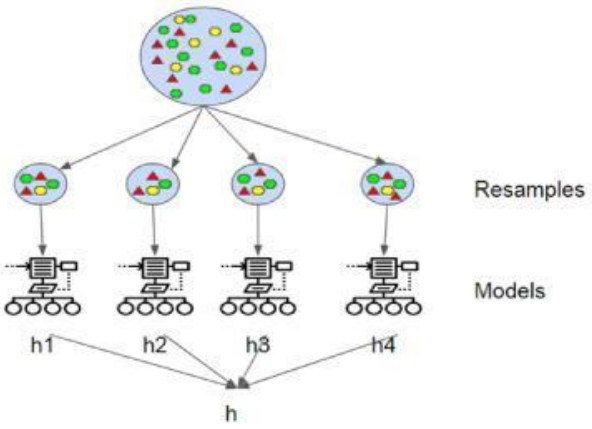
Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram



Bagging gets around this by creating its own variance amongst the data by sampling and replacing data while it tests multiple hypothesis(models). In turn, this reduces the noise by utilizing multiple samples that would most likely be made up of data with various attributes(median, average, etc).

Once each model has developed a hypothesis. The models use *voting for classification* or *averaging for regression*. This is where the "Aggregating" in "Bootstrap Aggregating" comes into play. Each hypothesis has the same weight as all the others. *When we later discuss boosting, this is one of the places the two methodologies differ.*



Essentially, all these models run at the same time, and vote on which hypothesis is the most accurate.

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

This helps to decrease variance i.e. reduce the overfit.

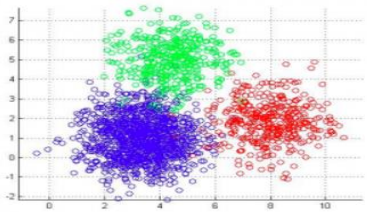
Advantages

- Bagging takes advantage of ensemble learning wherein multiple weak learners outperform a single strong learner.
- It helps reduce variance and thus helps us avoid overfitting.

Disadvantages

- There is a loss of interpretability of the model.
- There can possibly be a problem of high bias if not modeled properly.
- While bagging gives us more accuracy, it is computationally expensive and may not be desirable depending on the use case.

There are many bagging algorithms of which perhaps the most prominent would be Random Forest.

Q3a)  (5M)

i)K-Means can be used to solve **Unspervised** learning problems.
ii)In K-Means, K stands for __**No.of cluster**_____
iii)K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest **___Mean_____**
iv)--**KNN**- algorithm has similarity with K-Means?
v) The goal for K-Means cost function is to **_minimize_** squared error function where error function represents distance between data points and cluster centroid.

ii) How to decide the optimal number of K in the K mean algorithm?(5M)

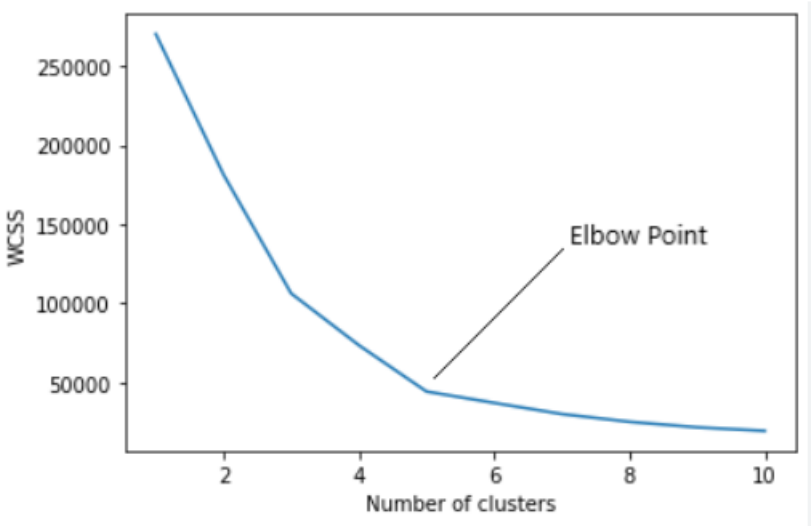
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

	he elbow method is used to determine the optimal number of clusters in k-means clustering. The elbow method plots the value of the cost function produced by different values of *k*. As you know, if *k* increases, average distortion will decrease, each cluster will have fewer constituent instances, and the instances will be closer to their respective centroids. However, the improvements in average distortion will decline as *k* increases. The value of *k* at which improvement in distortion declines the most is called the elbow, at which we should stop dividing the data into further clusters. Give sample example of K Vs WCSS 
b)	Test your skill about the Ensemble learning in terms of TRUE/False with justification. i)Ensemble learning can only be applied to supervised learning methods. A. True B. False Ans- **B Generally, we use ensemble technique for supervised learning algorithms. But, you can use anensemble for unsupervised learning algorithms also.**

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

ii) Ensembles will yield bad results when there is significant diversity among the models.

A.True
B. False

ANS- **B An ensemble is an art of combining a diverse set of learners (individual models) together to improvise on the stability and predictive power of the model. So, creating an ensemble of diverse models is a very important factor to achieve better results.**

iii)Ensemble of classifiers may or may not be more accurate than any of its individual model.

A. True

B. False

Ans- **True Usually, ensemble would improve the model, but it is not necessary. Hence, option A is correct.**

OR

What is Gaussian Mixture Model and When to use?(6+4)

Matematical formula of mixture model-
Data, Generative Story, Model, Log likehood
Show learning mixture model supervised ,unsupervised

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

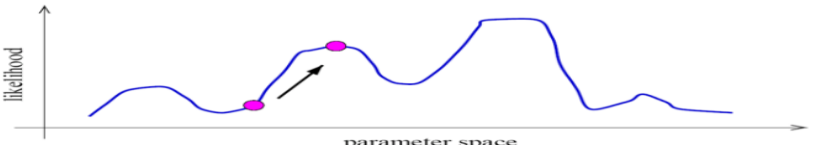
Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

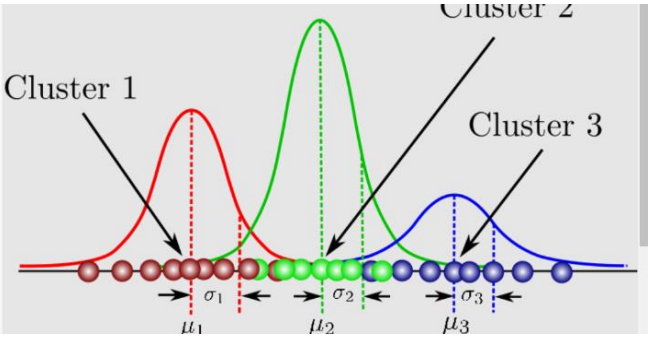
Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

Identifiability

- A mixture model induces a multi-modal likelihood.
- Hence gradient ascent can only find a local maximum.
- Mixture models are unidentifiable, since we can always switch the hidden labels without affecting the likelihood.
- Hence we should be careful in trying to interpret the "meaning" of latent variables.



Gaussian mixture models are very useful **when there are large datasets and it is difficult to find clusters**. This is where Gaussian mixture models help. It is able to find clusters of Gaussians more efficiently than other clustering algorithms such as k-means.



- Gaussian mixture models can be used for anomaly detection; by fitting a model to a dataset and then scoring new data points, it is possible to flag points that are significantly different from the rest of the data (i.e. outliers). This can be useful for identifying fraud or detecting errors in data collection.
- In the case of time series analysis, GMMs can be used to discover how volatility is related to trends and noise which can help predict future stock prices. One cluster could consist of a trend in the time series while another can have noise and volatility from other factors such as seasonality or external events which affect the stock price. In order to separate out these clusters, GMMs can be used because they provide a probability for each category instead of simply dividing the data into two parts such as that in the case of K-means.
- Another example is when there are different groups in a dataset and it's hard to label them as belonging to one group or another which makes it difficult for other machine learning algorithms such as the K-means clustering algorithm to separate out the

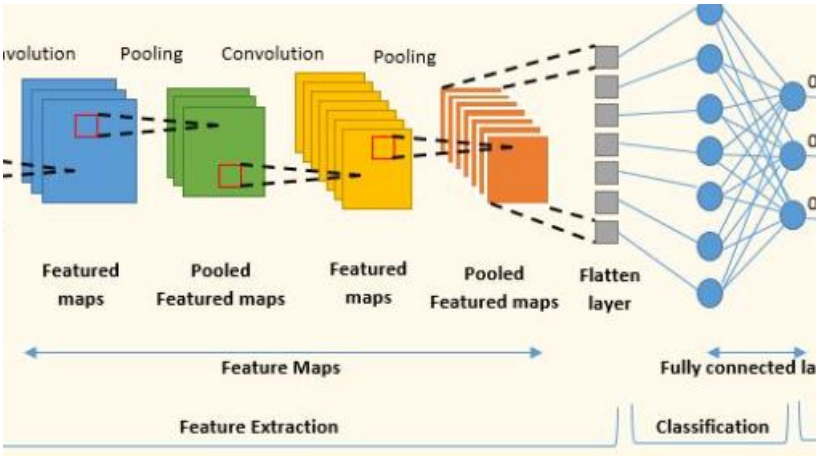
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

	data. GMMs can be used in this case because they find Gaussian mixture models that best describe each group and provide a probability for each cluster which is helpful when labeling clusters.
Q4 a)	Apply the concept of Kernel(3X3), stride, padding to generate feature map of image 5 X 5 for image classification. Show the details about CNN on 5x5 image. 1. Take the 5X 5 image 2Apply 3X3 Kernel and show output with detail calculation(4m) 2. Apply pooling and show the image (2M) 3. Give significance of padding and apply padding with kernel(2M) 4. Apply pooing and show the reduced 1D image and Fed to NN(2M) 5. Final diagram with details of each steps. Show i/p 5x5 image and size in each steps.(2M) 

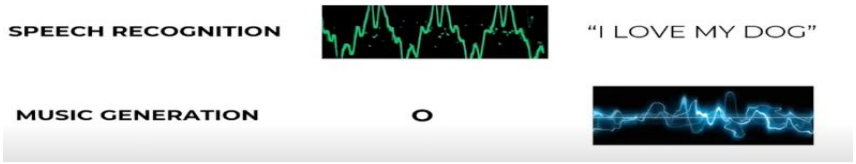
Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

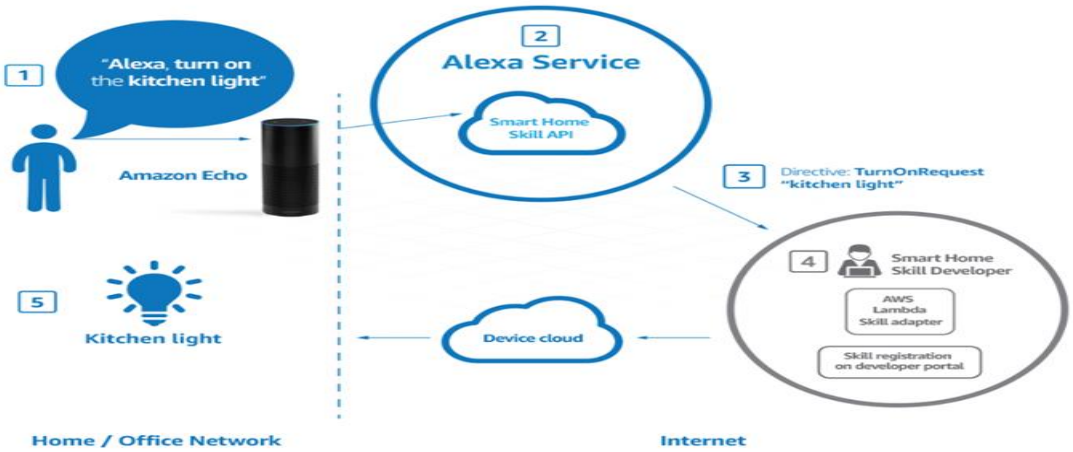


Identify the suitable neural network for the above applications. Discuss the architecture for the same.

Sol: Basic RNN architecture with methodical equations.(2M)

Show the Many to one RNN architecture for speech recognition(4M)
Architecture of one to many RNN with mathematical equations (4M)

Q5a) Alexa is built based on natural language processing (NLP), justify with a procedure of converting speech into words, sounds, and ideas.



Detail service by each block(8M)*Computational power requirement (2M)*

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

OR

Why NLP is the Next Frontier in AI for Enterprises

The power and practicality of artificial intelligence is largely undisputed. However, one of its longest-standing difficulties continues to be human language. A misspelled or misused word can frustratingly turn a chatbot conversation on its head, whereas a live service agent would have been able to prevent the mistake. Once computers can authentically understand human language, our business communications will completely transform. Machine learning (ML) and natural language processing (NLP) present us with a way to get there.

Importance

NLP is a branch of AI which enables computers to understand, write and speak languages just like humans. NLP has a number of use cases, including search engines, translation services, and chat or voice bots. Almost half of today's businesses use applications that are powered by NLP, and one in four businesses plan to begin using NLP technology within 12 months.

In most cases, NLP models rely on ML to comprehend written or spoken language. In this, the models learn much like humans do: using ML, they collect audio or written terms from humans, interpret the data, and can respond in the same language. Leveraging this process, some NLP-equipped AI algorithms can outperform humans in understanding the meaning of a text, or searching for the answer to a random question in over 100 languages. This level of proficiency has exciting implications for the business world. With NLP, searching for a product or service, or asking a brand a question can be quick and easy, no matter what language you speak or even whether you know how to phrase what you're looking for.

Use Cases

Chatbots are one of the most straightforward applications of NLP. With users' screen time up higher predominantly on messaging apps, it's critical that businesses leverage chatbots and AI assistants to be ready and able to respond to their customers around the clock.

NLP can help chatbots better understand customer inquiries and respond accordingly. This allows the bots to be strategic tools for a business' omnichannel communications by learning from previous

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

customer interactions to carry out a large volume of conversations in an efficient manner — all while reducing human errors and time consumption in customer service.

Additional use cases of NLP include:

Intelligent search: The majority of search engines work only with specific keywords rather than a true understanding of what a user is searching for. Intelligent search enabled by NLP allows users to ask questions, and the engine will pore over documents for an answer. Businesses should be looking to use their unstructured, textual data to better serve their customers via an intelligent search engine.

Automated topic and skill detection in chat: NLP-powered chatbots can make conversations more efficient by automating skills and detecting topics so they can hand off the conversation to the right agent if need be. For example, if a user is frustrated, NLP can detect that through their messages and respond accordingly to diffuse the situation and offer the customer the correct assistance.

Live agent support: Even when a chatbot is not leading a conversation with a customer, NLP can boost the efficiency of a conversation. NLP can create response suggestions based on the conversation and customer context so live agents don't need to type out individual answers. This can streamline efforts, allowing businesses to scale their customer service processes while ensuring all customers are receiving the attention that they need in a timely manner.

Benefits of NLP(in detail)

Large enterprises have the greatest potential to benefit from ML and AI due to their substantial pools of data, as well as the available budget for software that can process it. NLP can provide value to businesses of all sizes by allowing them to analyze and process unstructured data of any volume.

Due to its wide range of applications, NLP has a wide range of benefits for businesses who choose to incorporate it into their business strategy. These include (but are not limited to):

- **Boosting customer experience:**
- **Ensuring customers' needs are met:** .
- **Protecting employees and customers:**

Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

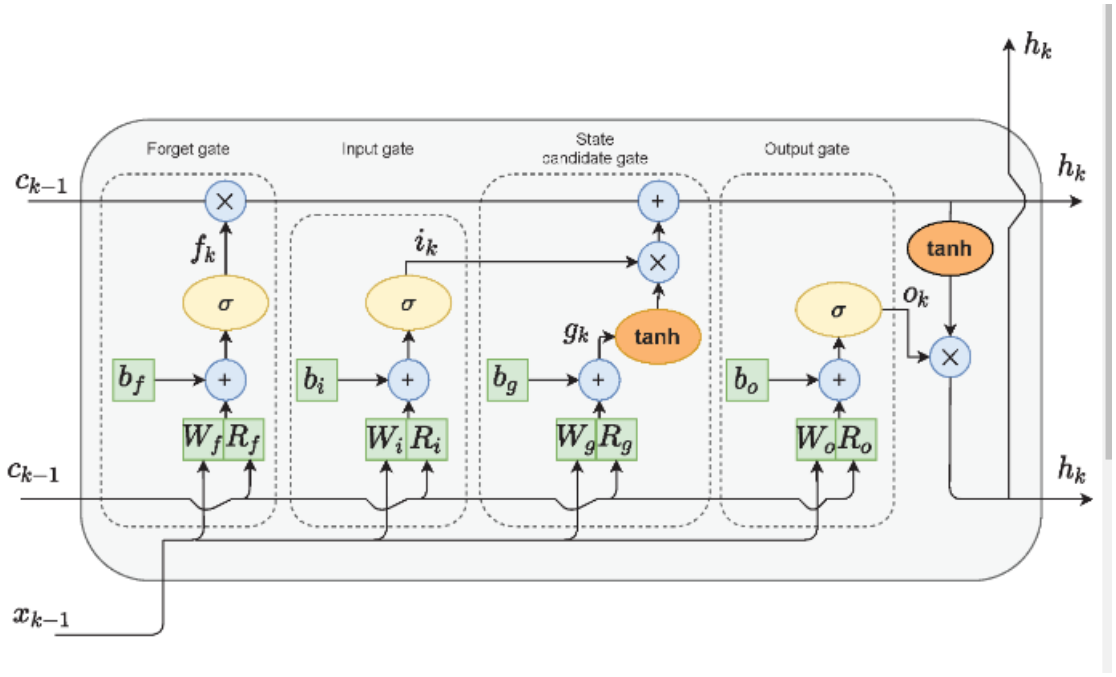
b) Design the suitable model to predict the word in the following sentence.

Komal eats dosa almost everyday it should not be hard to guess that his favorite cousin is -**South Indian**--------. Her sister Lata however is lover of sushi and paratha that means Lata's favorite Cuisine is -**Japanese.**

LSTM Model with all mathematical equations(4M)

Map the keyword with Forget , Input, and output Gate.

Need to show remove and add keywords in details for both the sentences (6M)



Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

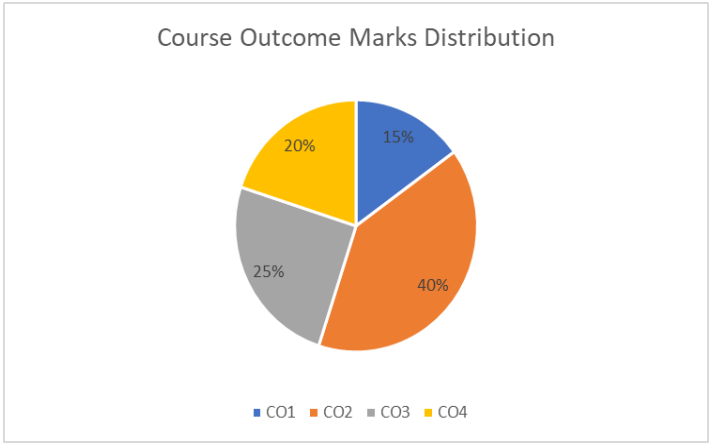
Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram

For official use only (not for students)

CO Number	CO statement	Marks allotted
CO1	Describe the basic concept and techniques of Machine learning	15
CO2	Evaluate supervised and unsupervised machine learning algorithm	40
CO3	Identify and analyze deep learning for various types of learning tasks in various domain	25
CO4	Apply knowledge representation, reasoning and machine learning techniques to real world problem	20

PI chart for percentage CO marks



Sardar Patel Institute of Technology
Bhavan's Campus, Munshi Nagar, Andheri (West), Mumbai-400058,
India
(Autonomous College Affiliated to University of Mumbai)

End Semester Examination

Max. Marks: 100
Class: B.E. **Duration: 3 Hr**
Course Code:EC433 **Semester: VII**
Branch: ETRX &EXTC

Name of the Course: Artificial Intelligence and Machine Learning

Instruction:
(1) All questions are compulsory
(2) Draw necessary diagram