

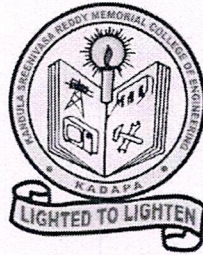
Kandula Srinivasa Reddy Memorial College of Engineering (Autonomous)

Kadapa-516003. AP

(Approved by AICTE, Affiliated to JNTUA, Ananthapuramu, Accredited by NAAC)

(An ISO 9001-2008 Certified Institution)

Department of Electronics and Communication Engineering



Certification Course

On

“Machine Learning for Beginners”

Resource Person : Dr. D. Arun Kumar

Course Coordinator: Smt. K.Divya Lakshmi

Smt. T. Uma Maheswari

Duration : 25-10-2021 to 10-11-2021



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - AUTONOMOUS)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



Lr./KSRMCE/ (Department of ECE)/2020-21

Date: 20-10-2021

To
The Principal
KSRM College of Engineering
Kadapa, AP.

Sub: KSRMCE - (Department of ECE) – Permission to conduct a certification course on “Machine Learning for Beginners” –Request– reg.

---***---

Respected Sir,

With reference to the cited, the Department of ECE is planning to conduct a certification course on “Machine Learning for Beginners” for All B.Tech V SEM students from 25-10-2021 to 10-11-2021. In this regard, we kindly request you to grant us permission to conduct a certification course. This is submitted for your kind perusal.

Thanking you sir,

Yours Faithfully,

Coordinators

Smt .T.Uma Maheswari

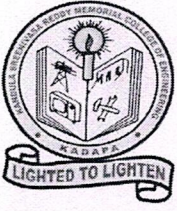
Smt .K. Divya Lakshmi

Cc:

To The Director for Information

To All Deans/HODs

Permitted
V. S. S. mm/5
PRINCIPAL
K.S.R.M. COLLEGE OF ENGINEERING
KADAPA-516005, (A.P.)



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - AUTONOMOUS)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



Date: 22-10-2021

Circular

All the B.Tech V SEM students are hereby informed that the department of ECE is going to conduct certification course on "Machine Learning for Beginners" from 25-10-2021 to 10-11-2021. Interested students may register their names with respective faculty members on or before 24-10-2021.

For any queries contact,

Coordinators

Smt. T. Uma Maheswari, Assistant Professor, ECE Dept.,

Smt.K. Divya Lakshmi , Assistant Professor, ECE Dept.,

HOD

Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003

Cc to:

The Management /Director / All Deans / All HODS/Staff / Students for information

The IQAC Cell for Documentation



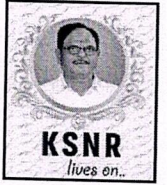
K.S.R.M. COLLEGE OF ENGINEERING

(UGC - AUTONOMOUS)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA,
Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



Department of Electronics & Communication Engineering
Certificate Course on Machine Learning for Beginners
Registration Form

S.No.	Roll.No.	Name of the Student	Branch	Sem	Signature
1	199Y1A0401	Akula Venkatesh	ECE	V	AV
2	199Y1A0402	Ambati Mouleswaraiah	ECE	V	Mouleswaraiah
3	199Y1A0403	ANGIATA KAVYA SREE	ECE	V	Kavya Sree
4	199Y1A0404	Animela Lavanya	ECE	V	Lavanya
5	199Y1A0405	Annemagani Mounika	ECE	V	Mounika
6	199Y1A0406	AVULA VARAJA	ECE	V	Varaja
7	199Y1A0407	Ayyaluri Venkata Pavan Kumar	ECE	V	Pavan Kumar
8	199Y1A0408	Balasamudram Arav Kumar	ECE	V	B. Arav
9	199Y1A0409	BANDI SAI BALAJI	ECE	V	B. Balaji
10	199Y1A0410	BANDI SANGEETHA	ECE	V	B. Sangeetha
11	199Y1A0411	BANDISEELA UDAY KUMAR	ECE	V	B. Uday Kumar
12	199Y1A0412	BANDIKULA SOORANATH	ECE	V	Sooranath
13	199Y1A0413	Bingimalla Venkatesh Kumar	ECE	V	B. Venkatesh Kumar
14	199Y1A0414	BOKKACAM SAIDEEPAK	ECE	V	Sai Deepak
15	199Y1A0415	BOSSI VENKATA SAINATH	ECE	V	B. Sainath
16	199Y1A0416	BUCHOPALLI MALATHI	ECE	V	Malathi
17	199Y1A0417	C. Jashwanth Varma	ECE	V	C. Jashwanth
18	199Y1A0418	CHAGANTI TEJESH KUMAR REDDY	ECE	V	T. Tejesh
19	199Y1A0419	Chalk Stephen Kumar	ECE	V	Stephen
20	199Y1A0420	CHAVALI RAJESH	ECE	V	C. Rajesh
21	199Y1A0421	Charva Sai Susmitha	ECE	V	S. Susmitha
22	199Y1A0422	Cheepati Varaprasad	ECE	V	V. Varaprasad
23	199Y1A0423	CHENNABOINA	ECE	V	C. Chennaboina

GURU PEEKSHITH

24	1994/A0424	CHEPPALI ANKAM	ECE	✓	Anakapali
25	1994/A0425	Chennai Ankam	ECE	✓	Chennai
26	1994/A0426	C. Geetha Priya	GCE	✓	Chennai
27	1994/A0427	C. Sai Dheeraj	ECE	✓	Chennai
28	1994/A0428	C. Sandeep Reddy	ECE	✓	Chennai
29	1994/A0429	C. Neetha Reddy	ECE	✓	Chennai
30	1994/A0430	C. Veeja Siva	ECE	✓	Chennai

K. P. Dnyaneshwari
Coordinator(s)

G. H. Anand
HOD

Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 083

Machine Learning for Beginners

30hrs

Course Objectives:

- Learn the basic fundamentals of Machine Learning.
- Understand Python software and relevant libraries.
- Gain knowledge to verify working of various Machine learning.
- Understand the use case of Machine Learning.

Course Outcomes:

The students will be able to:

- Understand and implement the most popular learning algorithms
- Perform feature selection and experimental set up on real tasks
- Evaluate multiple learning algorithms across several tasks.

Syllabus:

- Introduction to linear regression (and multivariate linear regression)
- Logistic Regression and regularization
- Practical aspects of implementation
- Decision trees and pruning, implementation of decision trees
- Support vector machines and making them work in practice
- Boosting - implementing different boosting methods with decision trees.
- Using the algorithms for several tasks - how to set up the problem, debug, select features and develop the learning algorithm.
- Unsupervised learning - k-means, PCA, hierarchical clustering.
- Implementing the clustering algorithms
- Parallelizing the learning algorithms
- Applications
- Choosing from multiple algorithms - What will work?

Text book:

1. <https://www.linkedin.com/learning/excel-data-analysis-forecasting/welcome>
2. A Course in Machine Learning, Hal Daume III (preprint available online)



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - AUTONOMOUS)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA,
Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



Department of Electronics & Communication Engineering

Certificate Course on Machine learning for beginners

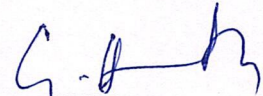
Schedule

S.No	Date	Time	Faculty	Topic
1	25-10-2021	3:00 PM to 4:00PM	Prof. A. Mohan	Inauguration
2	25-10-2021	4:00 PM to 5:00PM	Prof. A. Mohan	Introduction
3	26-10-2021	3:00 PM to 4:00PM	Prof. A. Mohan	Introduction of Machine learning,
4	26-10-2021	4:00 PM to 5:00PM	Prof. A. Mohan	Evaluation cross-validation
5	27-10-2021	3:00 PM to 4:00PM	Prof. A. Mohan	Basic definitions, types of learning
6	27-10-2021	4:00 PM to 5:00PM	Prof. A. Mohan	Basic definitions, types of learning
7	28-10-2021	3:00 PM to 4:00PM	Prof. A. Mohan	Hypothesis space and inductive bias evaluation
8	28-10-2021	4:00 PM to 5:00PM	Prof. A. Mohan	Hypothesis space and inductive bias evaluation
9	29-10-2021	3:00 PM to 4:00PM	Prof. A. Mohan	Cross-validation
10	29-10-2021	4:00 PM to 5:00PM	Prof. A. Mohan	Cross-validation

11	30-10-2021	3:00 PM to 4:00PM	Prof. A. Mohan	Linear regression on Machine learning.
12	30-10-2021	4:00 PM to 5:00PM	Prof. A. Mohan	Linear regression on Machine learning.
13	01/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Decision trees
14	01/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Decision trees
15	02/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Overfitting.
16	02/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Overfitting.
17	03/11/2021	2:00 PM to 3:00PM	Prof. A. Mohan	Instance based learning.
18	03/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Instance based learning
19	03/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Instance based learning
20	04/11/2021	2:00 PM to 3:00PM	Prof. A. Mohan	Feature reduction
21	04/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Feature reduction
22	04/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Feature reduction
23	05/11/2021	2:00 PM to 3:00PM	Prof. A. Mohan	Collaborative filtering based recommendation.
24	05/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Collaborative filtering based recommendation.
25	05/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Collaborative filtering based recommendation.

26	09/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Probability and bayes learning
27	09/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Probability and bayes learning
28	10/11/2021	2:00 PM to 3:00PM	Prof. A. Mohan	Bayes learning
29	10/11/2021	3:00 PM to 4:00PM	Prof. A. Mohan	Bayes learning
30	10/11/2021	4:00 PM to 5:00PM	Prof. A. Mohan	Valedictory

K. D. Nyabathi
Coordinator(s)


HOD

Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003



K.S.R.M. COLLEGE OF ENGINEERING (UGC - AUTONOMOUS)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.
An ISO 14001:2004 & 9001: 2015 Certified Institution

Department of Electronics & Communication Engineering

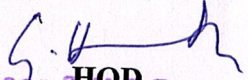
Certificate Course on Machine learning for beginners

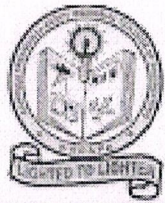
Attendance Sheet

S.No	Roll Num	Name of the Student	2	2	2	2	2	3	0	0	0	0	0	0	1
			5	6/	7	8	9	0	1	2	3	0	5	9	0
			/1	1	/	/	/	/	/	/	/	4	/	/	/
			0	0/	1	1	1	1	1	1	1	/	1	1	1
			/	2	0	0	0	0	1	1	1	1	1	1	1
			2	0	/	/	/	/	/	/	/	1	/	/	/
			0	2	2	2	2	2	2	2	2	/	2	2	2
			2	1	0	0	0	0	0	0	0	2	0	0	0
			1		2	2	2	2	2	2	2	0	2	2	2
					1	1	1	1	1	1	1	2	1	1	1
1	199Y1A0401	AKULA VENKATESH										1			
2	199Y1A0402	AMBATI MOULEESWARA REDDY	P	P	P	P	A	P	P	P	A	P	P	P	P
3	199Y1A0403	ANGAJALA KAVYA SREE (W)	P	P	P	A	A	P	P	P	P	P	P	P	P
4	199Y1A0404	ANIMELA LAVANYA (W)	P	P	P	P	P	A	A	P	P	P	P	P	P
5	199Y1A0405	ANNEMMAGARI MOUNIKA (W)	P	P	P	P	P	P	P	A	A	A	P	P	A
6	199Y1A0406	AVULA VANAJA (W)	P	P	A	A	P	P	P	P	P	P	P	P	P
7	199Y1A0407	AYYALURI VENKATA PAVAN KUMAR REDDY	P	P	P	P	A	A	P	P	P	A	A	P	P
8	199Y1A0408	BALASAMUDRAM AJAY KUMAR	P	P	P	A	P	A	P	P	P	A	P	P	P

9	199Y1A0409	BANDI SAI BALAJI	P	P	P	P	A	A	P	P	P	P	P	A	P
10	199Y1A0410	BANDI SANGEETHA (W)	P	P	P	P	P	P	P	P	A	P	P	P	P
11	199Y1A0411	BANDISEELA UDAYKUMAR	P	P	A	A	P	P	P	P	P	P	P	P	P
12	199Y1A0412	BARIVENKULA SREENATH	A	P	P	P	P	P	P	P	P	P	P	P	P
13	199Y1A0413	BINGIMALLA VENKATA THARUN KUMAR	P	P	P	P	A	A	P	P	P	P	P	P	P
14	199Y1A0414	BOKKASAM SAI DEEPAK	P	P	P	P	P	P	P	P	P	P	P	P	P
15	199Y1A0415	BOOSI VENKATA SAINATH REDDY	P	P	A	A	P	P	A	A	P	P	P	P	P
16	199Y1A0416	BUCHUPALLI MALATHI (W)	P	P	P	P	P	P	P	P	A	A	P	P	P
17	199Y1A0417	C JASHWANTH VARMA	P	P	A	P	P	P	P	P	P	P	P	P	P
18	199Y1A0418	CHAGANTI TEJESH KUMAR REDDY	P	P	P	A	P	P	P	A	P	P	A	P	P
19	199Y1A0419	CHALLA STEPHEN KUMAR	P	P	P	A	A	P	P	P	P	A	P	P	P
20	199Y1A0420	CHAVALI RAJESH	P	P	P	P	P	A	P	A	P	P	P	P	A
21	199Y1A0421	CHAVVA SAI SUSMITHA (W)	P	P	P	A	P	P	A	P	A	P	P	P	P
22	199Y1A0422	CHEEPATI VARAPRASAD REDDY	P	P	P	P	A	P	P	A	P	A	P	P	P
23	199Y1A0423	CHENNABOINA GURU DEEKSHITH	P	P	P	P	A	P	A	P	P	P	P	P	P
24	199Y1A0424	CHEPPALI ANKAIAH	P	P	P	P	P	P	P	P	P	P	P	P	P
25	199Y1A0425	CHERUVU SAI PRAKASH REDDY	P	P	P	P	P	P	P	P	P	P	P	P	P
26	199Y1A0426	CHILUMURU.GEETHA PRIYA (W)	P	A	P	P	P	P	A	P	P	P	P	A	P
27	199Y1A0427	CHINNAKOTLA SAI DHEERAJ	P	P	P	A	P	P	P	A	A	P	P	A	P
28	199Y1A0428	CHINNAULA SANDEEP REDDY	A	P	P	P	P	P	P	P	P	P	P	P	P
29	199Y1A0429	CHINNIREDDY NEETHA (W)	P	P	P	P	A	P	P	P	P	P	A	P	A
30	199Y1A0430	CHINTAKUNTA VEERA SIVA	P	P	P	P	P	P	A	P	P	P	P	P	P


Coordinator


Professor & HOD
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 518 003



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - Autonomous)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.



KSNR

Wise on.

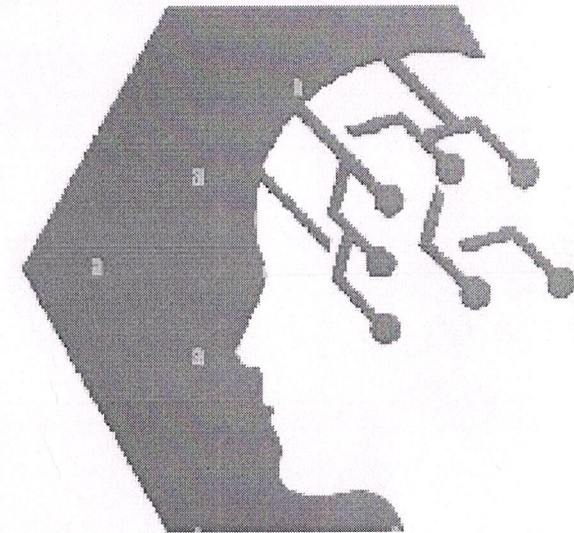
Department of ECE Certification Course on "Machine Learning for Beginners"



25-10-2021 to 10-11-2021



CRI Lab



Resource Persons

Prof. A. Mohan
Dept.of ECE,KSRMCE.

Coordinator

Smt.T. Uma Maheswari
Smt.K.Divya Lakshmi

Dr. G. Hemalatha
H.O.D

Dr. V.S.S. Murthy
Principal

Dr. Kandula Chandra Obul Reddy
Managing Director

Smt. K. Rajeswari
Correspondent Secretary,
Treasurer

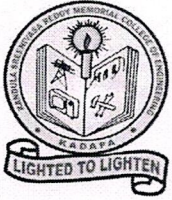
Sri K. Madan Mohan Reddy
Vice-Chairman

Sri K. Raja Mohan Reddy
Chairman

f @ t k [kasmceofficial](https://www.facebook.com/kasmceofficial)

www.kasmce.ac.in

8143731980, 8575697569



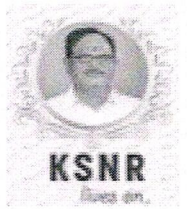
K.S.R.M. COLLEGE OF ENGINEERING

(UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



ACTIVITY REPORT

Certification Course

On

“Machine learning for beginners”

25th Oct 2021 to 10th Nov 2021

Target Group	:	V Sem ECE Students
Details of Participants	:	30 Students
Coordinators	:	Smt .T. Uma maheswari Asst.Prof, Dept. of ECE Smt .K. Divyalakshmi Asst. Prof, Dept. of ECE
Organizing Department	:	Department of Electronics and Communication Engineering
Venue	:	CRILab

Description:

Certification course on “Machine learning for beginners' ' was organized by Dept. of ECE from 25-10-2021 to 10-11-2021 . Prof .A. Mohan acted as Course instructor. Recent advances and Basics in Machine learning and Linear regression, Decision trees, overfitting , Instance based learning, Feature reduction, Collaborative filtering based recommendation , Probability and Bayes learning was explained.



/karmce.ac.in

Follow Us:



/karmceofficial

Photo :

 **K.S.R.M. COLLEGE OF ENGINEERING**
(UGC - Autonomous)
Kadapa, Andhra Pradesh, India - 516003
Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

 **KSNR**
Sri. en.

Department of ECE
Certification Course on
"Machine Learning for Beginners"

25-10-2021 to 10-11-2021 **CRI Lab**

Resource Persons
Prof. A. Mohan
Dept. of ECE, KSRMCE.

Coordinator
Smt. T. Uma Maheswari
Smt. K. Divya Lakshmi



Dr. G. Hemalatha H.O.D. Dr. V.S.S. Murthy Principal Dr. Kandula Chandra Obul Reddy Managing Director Smt. K. Rajeswari Correspondent Secretary, Treasurer Sri K. Madan Mohan Reddy Vice-Chairman Sri K. Raja Mohan Reddy Chairman

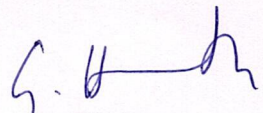
f @ /kasmceofficial www.kasmce.ac.in 8143731980, 8575697569




Smt. T. Umamaheswari

Smt.K.Divyalakshmi


Coordinators


HOD
Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003

 /kasmce.ac.in

Follow Us:



/kasmceofficial



KSRM COLLEGE OF ENGINEERING

UGC- AUTONOMOUS

Approved by AICTE, New Delhi, Affiliated to JNTUA, Ananthapuramu.
Kadapa, Andhra Pradesh, India- 516003.



KSNR
lives on..

CERTIFICATE OF COMPLETION

This is to certify that Mr./Ms. CHINNIREDDY NEETHA bearing roll number 199Y1A0429 has completed the course on **"Machine Learning for beginners"** organized by the department of **Electronics & Communication Engineering**, KSRM College of Engineering from 25-10-2021 to 10-11-2021.

HOD

Principal



KSRM COLLEGE OF ENGINEERING

UGC- AUTONOMOUS

Approved by AICTE, New Delhi, Affiliated to JNTUA, Ananthapuramu.
Kadapa, Andhra Pradesh, India- 516003.



KSNR
lives on..

CERTIFICATE OF COMPLETION

This is to certify that Mr./Ms. **BANDI SAI BALAJI** bearing roll number **199Y1A0409** has completed the course on "**Machine Learning for beginners**" organized by the department of **Electronics & Communication Engineering**, KSRM College of Engineering from **25-10-2021 to 10-11-2021**.

HOD

Principal



KSRM COLLEGE OF ENGINEERING

UGC- AUTONOMOUS
Approved by AICTE, New Delhi, Affiliated to JNTUA, Ananthapuramu.
Kadapa, Andhra Pradesh, India- 516003.



KSNR
lives on..

CERTIFICATE OF COMPLETION

This is to certify that Mr./Ms. CHEPPALI ANKAIAH bearing roll number 199Y1A0424 has completed the course on **"Machine Learning for beginners"** organized by the department of **Electronics & Communication Engineering**, KSRM College of Engineering from 25-10-2021 to 10-11-2021.

HOD

Principal



KSRM COLLEGE OF ENGINEERING

UGC- AUTONOMOUS
Approved by AICTE, New Delhi, Affiliated to JNTUA, Ananthapuramu.
Kadapa, Andhra Pradesh, India- 516003.



CERTIFICATE OF COMPLETION

This is to certify that Mr./Ms. CHALLA STEPHEN KUMAR
bearing roll number 19941A0419 has completed the course on
"Machine Learning for beginners" organized by the department of
Electronics & Communication Engineering, KSRM College of Engineering
from 25-10-2021 to 10-11-2021.

HOD

Principal

Feedback form on Certificate Course

Machine Learning for Beginners(25/10/2021 to 10/11/2021)

* Required

1. Roll Number *

2. Name of the Student *

3. B.Tech Semester *

Mark only one oval.

I SEM

II SEM

III SEM

IV SEM

V SEM

VI SEM

VII SEM

VIII SEM

4. Branch *

Mark only one oval.

Civil Enginerring

EEE

ME

ECE

CSE

AI&ML

5. Email ID *

6. Is the course content meet your exception. *

Mark only one oval.

Yes

No

7. Is the lecture sequence well planned. *

Mark only one oval.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

8. The contents of the course is explained with examples. *

Mark only one oval.

Strongly disagree

Disagree

Neutral

Agree

Strongly Agree

9. Is the level of course high. *

Mark only one oval.

Strongly disagree

Disagree

Neutral

Agree

Strongly Agree

10. Is the course exposed you to the new knowledge and practice. *

Mark only one oval.

Strongly disagree

Disagree

Neutral

Agree

Strongly Agree

11. Is the lecture clear and easy to understand. *

Mark only one oval.

Strongly disagree

Disagree

Neutral

Agree

Strongly agree

12. Rate the value of the course increasing your skills. *

Mark only one oval.

Strongly disagree

Disagree

Neutral

Agree

Strongly Agree

13. Any suggestions

This content is neither created nor endorsed by Google.

Google Forms



K.S.R.M. COLLEGE OF ENGINEERING (UGC - AUTONOMOUS)



Kadapa, Andhra Pradesh, India - 516003
Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.
Department of Electronics and Communication Engineering
Feedback Form

S.No.	Email address	Name of the student	Year & Semester	Branch	Roll Num	Is the course content met your expectation	Is the lecture sequence well planned	The contents of the course is explained with examples	Is the level of course high	Is the course exposed you to the new knowledge and practices	Is the lecturer clear and easy to understand	Rate the value of course in increasing your skills	Any issues
1	199Y1A0401@ksrmece.ac.in	AKULA VENKATESH	B.Tech Vsem	ECE	199Y1A0401	Yes	Yes	Agree	Agree	Strongly agree	4	5	Nothing
2	199Y1A0402@ksrmece.ac.in	AMBATI MOULEESWARA REDDY	B.Tech Vsem	ECE	199Y1A0402	Yes	Yes	Agree	Agree	Strongly agree	5	5	Nothing
3	199Y1A0403@ksrmece.ac.in	ANGAJALA KAVYA SREE (W)	B.Tech Vsem	ECE	199Y1A0403	Yes	Yes	Agree	Agree	Strongly agree	4	5	Good
4	199Y1A0404@ksrmece.ac.in	ANIMELA LAVANYA (W)	B.Tech Vsem	ECE	199Y1A0404	Yes	Yes	Agree	Agree	Strongly agree	5	5	nothing
5	199Y1A0405@ksrmece.ac.in	ANNEMMAGARI MOUNIKA (W)	B.Tech Vsem	ECE	199Y1A0405	Yes	Yes	Agree	Agree	Strongly agree	5	5	Good
6	199Y1A0406@ksrmece.ac.in	AVULA VANAJA (W)	B.Tech Vsem	ECE	199Y1A0406	Yes	Yes	Agree	Agree	Strongly agree	4	5	very good
7	199Y1A0407@ksrmece.ac.in	AYYALURI VENKATA PAVAN KUMAR REDDY	B.Tech Vsem	ECE	199Y1A0407	Yes	Yes	Strongly agree	Agree	Strongly agree	4	3	Nothing
8	199Y1A0408@ksrmece.ac.in	BALASAMUDRAM AJAY KUMAR	B.Tech Vsem	ECE	199Y1A0408	Yes	Yes	agree	Agree	Strongly agree	4	4	no
9	199Y1A0409@ksrmece.ac.in	BANDI SAI BALAJI	B.Tech Vsem	ECE	199Y1A0409	Yes	Yes	Strongly agree	Agree	Strongly agree	5	5	Nothing
10	199Y1A0410@ksrmece.ac.in	BANDI SANGEETHA (W)	B.Tech Vsem	ECE	199Y1A0410	Yes	Yes	Strongly agree	Agree	Strongly agree	5	5	Good
11	199Y1A0411@ksrmece.ac.in	BANDISEELA UDAYKUMAR	B.Tech Vsem	ECE	199Y1A0411	Yes	Yes	Agree	Agree	Strongly agree	5	4	Good


12	199Y1A0412@ksr ce.ac.in	BARIVENKULA SREENATH	B.Tech Vsem	ECE	199Y1A0412	Yes	Yes	agree	Agree	Strongly agree	5	5	Good
13	199Y1A0413@ksr ce.ac.in	BINGIMALLA VENKATA THARUN KUMAR	B.Tech Vsem	ECE	199Y1A0413	Yes	Yes	agree	Agree	Strongly agree	3	5	Good
14	199Y1A0414@ksr ce.ac.in	BOKKASAM SAI DEEPAK	B.Tech Vsem	ECE	199Y1A0414	Yes	Yes	agree	Agree	Strongly agree	5	4	very good
15	199Y1A0415@ksr ce.ac.in	BOOSI VENKATA SAINATH REDDY	B.Tech Vsem	ECE	199Y1A0415	Yes	Yes	agree	Agree	Strongly agree	4	4	very good
16	199Y1A0416@ksr ce.ac.in	BUCHUPALLI MALATHI (W)	B.Tech Vsem	ECE	199Y1A0416	Yes	Yes	agree	Agree	Strongly agree	5	4	very good
17	199Y1A0417@ksr ce.ac.in	C JASHWANTH VARMA	B.Tech Vsem	ECE	199Y1A0417	Yes	Yes	agree	Agree	Strongly agree	3	5	no
18	199Y1A0418@ksr ce.ac.in	CHAGANTI TEJESH KUMAR REDDY	B.Tech Vsem	ECE	199Y1A0418	Yes	Yes	agree	Agree	Strongly agree	4	5	nithing
19	199Y1A0419@ksr ce.ac.in	CHALLA STEPHEN KUMAR	B.Tech Vsem	ECE	199Y1A0419	Yes	Yes	Strongly agree	Agree	Strongly agree	4	5	Good
20	199Y1A0420@ksr ce.ac.in	CHAVALI RAJESH	B.Tech Vsem	ECE	199Y1A0420	Yes	Yes	Strongly agree	Agree	Strongly agree	4	4	Good
21	199Y1A0421@ksr ce.ac.in	CHAVVA SAI SUSMITHA (W)	B.Tech Vsem	ECE	199Y1A0421	Yes	Yes	Strongly agree	Agree	Strongly agree	4	3	Good
22	199Y1A0422@ksr ce.ac.in	CHEEPATI VARAPRASAD REDDY	B.Tech Vsem	ECE	199Y1A0422	Yes	Yes	agree	Agree	Strongly agree	4	4	Good
23	199Y1A0423@ksr ce.ac.in	CHENNABOINA GURU DEEKSHITH	B.Tech Vsem	ECE	199Y1A0423	Yes	Yes	agree	Agree	Strongly agree	5	4	Good
24	199Y1A0424@ksr ce.ac.in	CHEPPALI ANKAIAH	B.Tech Vsem	ECE	199Y1A0424	Yes	Yes	Strongly agree	Agree	Strongly agree	5	4	Good
25	199Y1A0425@ksr ce.ac.in	CHERUVU SAI PRAKASH REDDY	B.Tech Vsem	ECE	199Y1A0425	Yes	Yes	agree	Agree	Strongly agree	5	5	Good
26	199Y1A0426@ksr ce.ac.in	CHILUMURU.GEETHA PRIYA (W)	B.Tech Vsem	ECE	199Y1A0426	Yes	Yes	agree	Agree	Strongly agree	5	5	Nothing
27	199Y1A0427@ksr ce.ac.in	CHINNAKOTLA SAI DHEERAJ	B.Tech Vsem	ECE	199Y1A0427	Yes	Yes	agree	Agree	Strongly agree	5	5	no
28	199Y1A0428@ksr ce.ac.in	CHINNAULA SANDEEP REDDY	B.Tech Vsem	ECE	199Y1A0428	Yes	Yes	agree	Agree	Strongly agree	3	4	no
29	199Y1A0429@ksr ce.ac.in	CHINNI REDDY NEETHA (W)	B.Tech Vsem	ECE	199Y1A0429	Yes	Yes	Strongly agree	Agree	Strongly agree	3	4	no
30	199Y1A0430@ksr ce.ac.in	CHINTAKUNTA VEERA SIVA	B.Tech Vsem	ECE	199Y1A0430	Yes	Yes	Strongly agree	Agree	Strongly agree		5	no

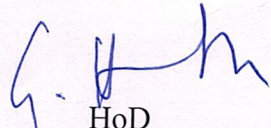
Professor & H.O.D.
 Department of E.C.E.
 K.S.R.M. College of Engineering
 KADAPA - 516 003

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
VALUE ADDED/CERTIFICATE COURSE ON
MACHINE LEARNING FOR BEGINNERS FROM 25/10/2021 to 10/11/2021
AWARD LIST

S.No	Roll Number	Name of the Student	Marks Obtained
1.	199Y1A0469	KETHIREDDY PRASANTHI	15
2.	199Y1A0470	KODURI LAKSHMI NARAYANA	12
3.	199Y1A0471	KOMPALA SAI CHARAN	15
4.	199Y1A0472	KONDA PRATHYUSHA	16
5.	199Y1A0473	KONDAMUGARI EESHITHA RACHANA RAVINDRA	14
6.	199Y1A0474	KONDURU SUMITHRA	15
7.	199Y1A0475	KOTHAMADDI NEHA	15
8.	199Y1A0476	KOTIREDDY CHANDRA SEKHAR REDDY	16
9.	199Y1A0477	KOTTHA REDDY ARCHANA	11
10.	199Y1A0478	KUNDELLA DEEVAN KUMAR	05
11.	199Y1A0479	KUNDHARAPU VENKATESH	12
12.	199Y1A0480	KURUBA BYALLA YERRISWAMY	13
13.	199Y1A0481	KURUBA LAKSHMAN	06
14.	199Y1A0482	KURUVA ANJI	09
15.	199Y1A0483	KURUVA LAKSHMANNA	15
16.	199Y1A0484	KURUVA SAI PRAKASH	15
17.	199Y1A0485	KUTEDDULA JASWANTH REDDY	14
18.	199Y1A0486	LAKKIREDDYPALLI RUPESH BABU	13
19.	199Y1A0487	M SAI VARDHAN NAIDU	12
20.	199Y1A0488	MACHIREDDY VENKATA SAI NATH REDDY	12

21.	199Y1A0489	MACHUNURU SREE VANI	13
22.	199Y1A0490	MADDIREDDY PRAKASH REDDY	12
23.	199Y1A0491	MALISSETTY GIRI BABU	14
24.	199Y1A0492	MALISSETTY HARIKA	14
25.	199Y1A0493	MEDA SUDHA RANI	15
26.	199Y1A0495	MEKA CHANDRA SEK HAR	09
27.	199Y1A0496	MITTA VENKATA MANISHA	14
28.	199Y1A0497	MOOLA GANGADHAR REDDY	13
29.	199Y1A0498	MOPURI GOWTHAM	11
30.	199Y1A0499	MOPURU NAVANEETHA	12


Coordinator


HoD
Professor & H.O.D.
Department of E.C.E.
R.S.R.M. College of Engineering
KADAPA - 518 082.

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATIONS OF ENGINEERING
VALUE ADDED /CERTIFICATE COURSE ON
MACHINE LEARNING FOR BEGINNERS FROM 25/10/2021 to 10/11/2021
ASSESSMENT TEST

Roll Number: _____ **Name of the Student:** _____

Time: 20 Min

(Objective Questions)

Max.Marks: 20

Note: Answer the following Questions and each question carries **one** mark.

1. What is true about Machine Learning? []
A. Machine Learning (ML) is that field of computer science
B. ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method.
C. The main focus of ML is to allow computer systems learn from experience without being Explicitly programmed or human intervention.
D. All of the above
2. ML is a field of AI consisting of learning algorithms that? []
A. Improve their performance B. At executing some task
C. Over time with experience D. All of the above
3. $p \rightarrow 0q$ is not a? []
A. hack clause B. horn clause C. structural clause D. system clause
4. The action _____ of a robot arm specify to Place block A on block B. []
A. STACK(A,B) B. LIST(A,B) C. QUEUE(A,B) D. ARRAY(A,B)
5. A _____ begins by hypothesizing a sentence (the symbol S) and successively predicting lower level constituents until individual preterminal symbols are written. []
A. bottow-up parser B. top parser C. top-down parser D. bottom parser
6. A model of language consists of the categories which does not include _____. []
A. System Unit B. structural units. C. data units D. empirical units
7. Different learning methods does not include? []
A. Introduction B. Analogy C. Deduction D. Memorization

8. The model will be trained with data in one single batch is known as ? []

- A. Batch learning B. Offline learning C. Both A and B D. None of the above

9. Which of the following are ML methods? []

- A. based on human supervision B. supervised Learning
C. semi-reinforcement Learning D. All of the above

10. In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called ? []

- A. mini-batches B. optimized parameters C. hyper parameters D. super parameters

11. which of the following is a widely used and effective machine learning algorithm based on the idea of bagging? []

- A. Decision Tree B. Regression C. Classification D. Random Forest

12. To find the minimum or the maximum of a function, we set the gradient to zero because:

- A. The value of the gradient at extrema of a function is always zero []
B. Depends on the type of problem C. Both A and B D. None of the above

13. Which of the following is a disadvantage of decision trees? []

- A. Factor analysis B. Decision trees are robust to outliers
C. Decision trees are prone to be overfit D. None of the above

14. How do you handle missing or corrupted data in a dataset? []

- A. Drop missing rows or columns B. Replace missing values with mean/median/mode
C. Assign a unique category to missing values D. All of the above

15. When performing regression or classification, which of the following is the correct way to preprocess the data? []

- A. Normalize the data -> PCA -> training B. PCA -> normalize PCA output -> training
C. Normalize the data -> PCA -> normalize PCA output -> training D. None of the above

16. Which of the following statements about regularization is not correct? []

- A. Using too large a value of lambda can cause your hypothesis to underfit the data.
B. Using too large a value of lambda can cause your hypothesis to overfit the data
C. Using a very large value of lambda cannot hurt the performance of your hypothesis.
D. None of the above

17. Which of the following techniques can not be used for normalization in text mining? []

- A. Stemming B. Lemmatization C. Stop Word Removal D. None of the above

18. In which of the following cases will K-means clustering fail to give good results? []

- 1) Data points with outliers
- 2) Data points with different densities
- 3) Data points with nonconvex shapes

- A. 1 and 2 B. 2 and 3 C. 1 and 3 D. All of the above

19. Which of the following is a reasonable way to select the number of principal components "k"? []

- A. Choose k to be the smallest value so that at least 99% of the variance is retained.
- B. Choose k to be 99% of m ($k = 0.99 * m$, rounded to the nearest integer).
- C. Choose k to be the largest value so that 99% of the variance is retained.
- D. Use the elbow method.

20. What is a sentence parser typically used for? []

- A. It is used to parse sentences to check if they are utf-8 compliant.
- B. It is used to parse sentences to derive their most likely syntax tree structures.
- C. It is used to parse sentences to assign POS tags to all tokens.
- D. It is used to check if sentences can be parsed into meaningful tokens.

12/20

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATIONS OF ENGINEERING
VALUE ADDED /CERTIFICATE COURSE ON
MACHINE LEARNING FOR BEGINNERS FROM 25/10/2021 to 10/11/2021

ASSESSMENT TEST

Roll Number: 199Y1A0479 **Name of the Student:** K. Venkatesh

Time: 20 Min **(Objective Questions)** **Max.Marks: 20**

Note: Answer the following Questions and each question carries **one** mark.

1. What is true about Machine Learning?

[D]

- A. Machine Learning (ML) is that field of computer science
- B. ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method.
- C. The main focus of ML is to allow computer systems learn from experience without being Explicitly programmed or human intervention.
- D. All of the above

2. ML is a field of AI consisting of learning algorithms that?

[D]

- A. Improve their performance
- B. At executing some task
- C. Over time with experience
- D. All of the above

3. $p \rightarrow 0q$ is not a?

[B]

- A. hack clause
- B. horn clause
- C. structural clause
- D. system clause

4. The action _____ of a robot arm specify to Place block A on block B.

[B]

- A. STACK(A,B)
- B. LIST(A,B)
- C. QUEUE(A,B)
- D. ARRAY(A,B)

5. A _____ begins by hypothesizing a sentence (the symbol S) and successively predicting lower level constituents until individual preterminal symbols are written.

[B]

- A. bottow-up parser
- B. top parser
- C. top-down parser
- D. bottom parser

6. A model of language consists of the categories which does not include _____.

[A]

- A. System Unit
- B. structural units.
- C. data units
- D. empirical units

7. Different learning methods does not include?

[A]

- A. Introduction
- B. Analogy
- C. Deduction
- D. Memorization

8. The model will be trained with data in one single batch is known as ? [B]
A. Batch learning B. Offline learning C. Both A and B D. None of the above

9. Which of the following are ML methods? [D]
A. based on human supervision B. supervised Learning
C. semi-reinforcement Learning D. All of the above

10. In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called ? [C]
A. mini-batches B. optimized parameters C. hyper parameters D. super parameters

11. which of the following is a widely used and effective machine learning algorithm based on the idea of bagging? [C]
A. Decision Tree B. Regression C. Classification D. Random Forest

12. To find the minimum or the maximum of a function, we set the gradient to zero because:
A. The value of the gradient at extrema of a function is always zero [A]
B. Depends on the type of problem C. Both A and B D. None of the above

13. Which of the following is a disadvantage of decision trees? [B]
A. Factor analysis B. Decision trees are robust to outliers
C. Decision trees are prone to be overfit D. None of the above

14. How do you handle missing or corrupted data in a dataset? [D]
A. Drop missing rows or columns B. Replace missing values with mean/median/mode
C. Assign a unique category to missing values D. All of the above

15. When performing regression or classification, which of the following is the correct way to preprocess the data? [C]
A. Normalize the data -> PCA -> training B. PCA -> normalize PCA output -> training
C. Normalize the data -> PCA -> normalize PCA output -> training D. None of the above

16. Which of the following statements about regularization is not correct? [B]
A. Using too large a value of lambda can cause your hypothesis to underfit the data.
B. Using too large a value of lambda can cause your hypothesis to overfit the data
C. Using a very large value of lambda cannot hurt the performance of your hypothesis.
D. None of the above

17. Which of the following techniques can not be used for normalization in text mining? [C]

- A. Stemming B. Lemmatization C. Stop Word Removal D. None of the above

18. In which of the following cases will K-means clustering fail to give good results? [C]

- 1) Data points with outliers
- 2) Data points with different densities
- 3) Data points with nonconvex shapes

- A. 1 and 2 B. 2 and 3 C. 1 and 3 D. All of the above

19. Which of the following is a reasonable way to select the number of principal components "k"? [B]

- A. Choose k to be the smallest value so that at least 99% of the variance is retained.
- B. Choose k to be 99% of m ($k = 0.99 * m$, rounded to the nearest integer).
- C. Choose k to be the largest value so that 99% of the variance is retained.
- D. Use the elbow method.

20. What is a sentence parser typically used for?

- A. It is used to parse sentences to check if they are utf-8 compliant.
- B. It is used to parse sentences to derive their most likely syntax tree structures.
- C. It is used to parse sentences to assign POS tags to all tokens.
- D. It is used to check if sentences can be parsed into meaningful tokens.

[B]

15/10

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATIONS OF ENGINEERING
VALUE ADDED /CERTIFICATE COURSE ON
MACHINE LEARNING FOR BEGINNERS FROM 25/10/2021 to 10/11/2021

ASSESSMENT TEST

Roll Number: 19991A0484 Name of the Student: KURUVA Sai prakash

Time: 20 Min (Objective Questions) **Max.Marks: 20**

Note: Answer the following Questions and each question carries **one** mark.

1. What is true about Machine Learning?

[D]

- A. Machine Learning (ML) is that field of computer science
- B. ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method.
- C. The main focus of ML is to allow computer systems learn from experience without being Explicitly programmed or human intervention.
- D. All of the above

2. ML is a field of AI consisting of learning algorithms that?

[D]

- A. Improve their performance
- B. At executing some task
- C. Over time with experience
- D. All of the above

3. $p \rightarrow 0q$ is not a?

[B]

- A. hack clause
- B. horn clause
- C. structural clause
- D. system clause

4. The action _____ of a robot arm specify to Place block A on block B.

[A]

- A. STACK(A,B)
- B. LIST(A,B)
- C. QUEUE(A,B)
- D. ARRAY(A,B)

5. A _____ begins by hypothesizing a sentence (the symbol S) and successively predicting lower level constituents until individual preterminal symbols are written.

[B]

- A. bottom-up parser
- B. top parser
- C. top-down parser
- D. bottom parser

6. A model of language consists of the categories which does not include _____.

[A]

- A. System Unit
- B. structural units.
- C. data units
- D. empirical units

7. Different learning methods does not include?

[A]

- A. Introduction
- B. Analogy
- C. Deduction
- D. Memorization

8. The model will be trained with data in one single batch is known as ? [A]
A. Batch learning B. Offline learning C. Both A and B D. None of the above
9. Which of the following are ML methods? [C]
A. based on human supervision B. supervised Learning
C. semi-reinforcement Learning D. All of the above
10. In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called ? [B]
A. mini-batches B. optimized parameters C. hyper parameters D. super parameters
11. which of the following is a widely used and effective machine learning algorithm based on the idea of bagging? [D]
A. Decision Tree B. Regression C. Classification D. Random Forest
12. To find the minimum or the maximum of a function, we set the gradient to zero because:
A. The value of the gradient at extrema of a function is always zero [A]
B. Depends on the type of problem C. Both A and B D. None of the above
13. Which of the following is a disadvantage of decision trees? [C]
A. Factor analysis B. Decision trees are robust to outliers
C. Decision trees are prone to be overfit D. None of the above
14. How do you handle missing or corrupted data in a dataset? [C]
A. Drop missing rows or columns B. Replace missing values with mean/median/mode
C. Assign a unique category to missing values D. All of the above
15. When performing regression or classification, which of the following is the correct way to preprocess the data? [C]
A. Normalize the data -> PCA -> training B. PCA -> normalize PCA output -> training
C. Normalize the data -> PCA -> normalize PCA output -> training D. None of the above
16. Which of the following statements about regularization is not correct? [A]
A. Using too large a value of lambda can cause your hypothesis to underfit the data.
B. Using too large a value of lambda can cause your hypothesis to overfit the data
C. Using a very large value of lambda cannot hurt the performance of your hypothesis.
D. None of the above

17. Which of the following techniques can not be used for normalization in text mining? [D]

- A. Stemming B. Lemmatization C. Stop Word Removal D. None of the above

18. In which of the following cases will K-means clustering fail to give good results? [A]

- 1) Data points with outliers
- 2) Data points with different densities
- 3) Data points with nonconvex shapes

- A. 1 and 2 B. 2 and 3 C. 1 and 3 D. All of the above

19. Which of the following is a reasonable way to select the number of principal components "k"?

- [B]
- A. Choose k to be the smallest value so that at least 99% of the variance is retained.
 - B. Choose k to be 99% of m ($k = 0.99 * m$, rounded to the nearest integer).
 - C. Choose k to be the largest value so that 99% of the variance is retained.
 - D. Use the elbow method.

20. What is a sentence parser typically used for?

- [B]
- A. It is used to parse sentences to check if they are utf-8 compliant.
 - B. It is used to parse sentences to derive their most likely syntax tree structures.
 - C. It is used to parse sentences to assign POS tags to all tokens.
 - D. It is used to check if sentences can be parsed into meaningful tokens.

14/20

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATIONS OF ENGINEERING
VALUE ADDED /CERTIFICATE COURSE ON
MACHINE LEARNING FOR BEGINNERS FROM 25/10/2021 to 10/11/2021

ASSESSMENT TEST

Roll Number: 199VIA0496 **Name of the Student:** M. Venkata Manisha

Time: 20 Min **(Objective Questions)** **Max.Marks: 20**

Note: Answer the following Questions and each question carries **one** mark.

1. What is true about Machine Learning?

[D]

- A. Machine Learning (ML) is that field of computer science
- B. ML is a type of artificial intelligence that extract patterns out of raw data by using an algorithm or method.
- C. The main focus of ML is to allow computer systems learn from experience without being Explicitly programmed or human intervention.
- D. All of the above

2. ML is a field of AI consisting of learning algorithms that?

[D]

- A. Improve their performance
- B. At executing some task
- C. Over time with experience
- D. All of the above

3. $p \rightarrow 0q$ is not a?

[B]

- A. hack clause
- B. horn clause
- C. structural clause
- D. system clause

4. The action _____ of a robot arm specify to Place block A on block B.

[A]

- A. STACK(A,B)
- B. LIST(A,B)
- C. QUEUE(A,B)
- D. ARRAY(A,B)

5. A _____ begins by hypothesizing a sentence (the symbol S) and successively predicting lower level constituents until individual preterminal symbols are written.

[B]

- A. bottow-up parser
- B. top parser
- C. top-down parser
- D. bottom parser

6. A model of language consists of the categories which does not include _____.

[A]

- A. System Unit
- B. structural units.
- C. data units
- D. empirical units

7. Different learning methods does not include?

[A]

- A. Introduction
- B. Analogy
- C. Deduction
- D. Memorization

3.5/11

8. The model will be trained with data in one single batch is known as ? [A]

- A. Batch learning B. Offline learning C. Both A and B D. None of the above

9. Which of the following are ML methods? [C]

- A. based on human supervision B. supervised Learning
C. semi-reinforcement Learning D. All of the above

10. In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called ? [B]

- A. mini-batches B. optimized parameters C. hyper parameters D. super parameters

11. which of the following is a widely used and effective machine learning algorithm based on the idea of bagging? [D]

- A. Decision Tree B. Regression C. Classification D. Random Forest

12. To find the minimum or the maximum of a function, we set the gradient to zero because:

- A. The value of the gradient at extrema of a function is always zero [A]
B. Depends on the type of problem C. Both A and B D. None of the above

13. Which of the following is a disadvantage of decision trees? [C]

- A. Factor analysis B. Decision trees are robust to outliers
C. Decision trees are prone to be overfit D. None of the above

14. How do you handle missing or corrupted data in a dataset? [C]

- A. Drop missing rows or columns B. Replace missing values with mean/median/mode
C. Assign a unique category to missing values D. All of the above

15. When performing regression or classification, which of the following is the correct way to preprocess the data? [C]

- A. Normalize the data -> PCA -> training B. PCA -> normalize PCA output -> training
C. Normalize the data -> PCA -> normalize PCA output -> training D. None of the above

16. Which of the following statements about regularization is not correct? [A]

- A. Using too large a value of lambda can cause your hypothesis to underfit the data.
B. Using too large a value of lambda can cause your hypothesis to overfit the data
C. Using a very large value of lambda cannot hurt the performance of your hypothesis.
D. None of the above

17. Which of the following techniques can not be used for normalization in text mining? [~~D~~]

- A. Stemming B. Lemmatization C. Stop Word Removal D. None of the above

18. In which of the following cases will K-means clustering fail to give good results? [~~A~~]

- 1) Data points with outliers
- 2) Data points with different densities
- 3) Data points with nonconvex shapes

- A. 1 and 2 B. 2 and 3 C. 1 and 3 D. All of the above

19. Which of the following is a reasonable way to select the number of principal components "k"?

[~~B~~]

- A. Choose k to be the smallest value so that at least 99% of the variance is retained.
- B. Choose k to be 99% of m ($k = 0.99 * m$, rounded to the nearest integer).
- C. Choose k to be the largest value so that 99% of the variance is retained.
- D. Use the elbow method.

20. What is a sentence parser typically used for?

[~~B~~]

- A. It is used to parse sentences to check if they are utf-8 compliant.
- B. It is used to parse sentences to derive their most likely syntax tree structures.
- C. It is used to parse sentences to assign POS tags to all tokens.
- D. It is used to check if sentences can be parsed into meaningful tokens.

Machine Learning for Beginners

What is Machine Learning?

Arthur Samuel coined the term Machine Learning in the year 1959. He was a pioneer in Artificial Intelligence and computer gaming, and defined Machine Learning as "Field of study that gives computers the capability to learn without being explicitly programmed". In simple terms, Machine Learning is an application of Artificial Intelligence (AI) which enables a program (software) to learn from the experiences and improve their self at a task without being explicitly programmed. For example, how would you write a program that can identify fruits based on their various properties, such as colour, shape, size or any other property?

One approach is to hardcode everything, make some rules and use them to identify the fruits. This may seem the only way and work but one can never make perfect rules that apply on all cases. This problem can be easily solved using machine learning without any rules which makes it more robust and practical. You will see how we will use machine learning to do this task in the coming sections.

Thus, we can say that Machine Learning is the study of making machines more human-like in their behaviour and decision making by giving them the ability to learn with minimum human intervention, i.e., no explicit programming. Now the question arises, how can a program attain any experience and from where does it learn? The answer is data. Data is also called the fuel for Machine Learning and we can safely say that there is no machine learning without data.

You may be wondering that the term Machine Learning has been introduced in 1959 which is a long way back, then why haven't there been any mention of it till recent years? You may want to note that Machine Learning needs a huge computational power, a lot of data and devices which are capable of storing such vast data. We have only recently reached a point where we now have all these requirements and can practice Machine Learning.

How is it different from traditional programming?

Are you wondering how is Machine Learning different from traditional programming? Well, in traditional programming, we would feed the input data and a well written and tested program into a machine to generate output. When it comes to machine learning, input data along with the output associated with the data is fed into the machine during

the learning phase, and it works out a program for itself. To understand this better, refer to the illustration below:

Don't worry if you are not able to understand this completely, in the coming sections you will get a better understanding. You may want to come back to this figure once we discuss the steps that are involved in machine learning to clear all your doubts.

Why do we need Machine Learning?

Machine Learning today has all the attention it needs. Machine Learning can automate many tasks, especially the ones that only humans can perform with their innate intelligence. Replicating this intelligence to machines can be achieved only with the help of machine learning.

With the help of Machine Learning, businesses can automate routine tasks. It also helps in automating and quickly create models for data analysis. Various industries depend on vast quantities of data to optimize their operations and make intelligent decisions. Machine Learning helps in creating models that can process and analyze large amounts of complex data to deliver accurate results. These models are precise and scalable and function with less turnaround time. By building such precise Machine Learning models, businesses can leverage profitable opportunities and avoid unknown risks.

Image recognition, text generation, and many other use-cases are finding applications in the real world. This is increasing the scope for machine learning experts to shine as a sought after professionals.

How Does Machine Learning Work?

A machine learning model learns from the historical data fed to it and then builds prediction algorithms to predict the output for the new set of data that comes in as input to the system. The accuracy of these models would depend on the quality and amount of input data. A large amount of data will help build a better model which predicts the output more accurately.

Suppose we have a complex problem at hand that requires to perform some predictions. Now, instead of writing a code, this problem could be solved by feeding the given data to generic machine learning algorithms. With the help of these algorithms, the machine will develop logic and predict the output. Machine learning has transformed the way we approach business and social problems. Below is a diagram that briefly explains the working of a machine learning model/ algorithm. Our way of thinking about the problem. The below block diagram explains the working of Machine Learning algorithm:

History of Machine Learning

Nowadays, we can see some amazing applications of ML such as in self-driving cars, Natural Language Processing and many more. But Machine learning has been here for over 70 years now. It all started in 1943, when neurophysiologist Warren McCulloch and mathematician Walter Pitts wrote a paper about neurons, and how they work. They decided to create a model of this using an electrical circuit, and therefore, the neural network was born.

In 1950, Alan Turing created the "Turing Test" to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human. In 1952, Arthur Samuel wrote the first computer learning program. The program was the game of checkers, and the IBM computer improved at the game the more it played, studying which moves made up winning strategies and incorporating those moves into its program.

Just after a few years, in 1957, Frank Rosenblatt designed the first neural network for computers (the perceptron), which simulates the thought processes of the human brain. Later, in 1967, the "nearest neighbour" algorithm was written, allowing computers to begin using very basic pattern recognition. This could be used to map a route for travelling salesmen, starting at a random city but ensuring they visit all cities during a short tour.

But we can say that in the 1990s we saw a big change. Now work on machine learning shifted from a knowledge-driven approach to a data-driven approach. Scientists began to create programs for computers to analyze large amounts of data and draw conclusions or "learn" from the results.

In 1997, IBM's Deep Blue became the first computer chess-playing system to beat a reigning world chess champion. Deep Blue used the computing power in the 1990s to perform large-scale searches of potential moves and select the best move. Just a decade before this, in 2006, Geoffrey Hinton created the term "deep learning" to explain new algorithms that help computers distinguish objects and text in images and videos.

Machine Learning at Present

The year 2012 saw the publication of an influential research paper by Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever, describing a model that can dramatically reduce the error rate in image recognition systems. Meanwhile, Google's X Lab developed a machine learning algorithm capable of autonomously browsing YouTube videos to identify the videos that contain cats. In 2016 AlphaGo (created by researchers at Google DeepMind to play the ancient Chinese game of Go) won four out of five matches against Lee Sedol, who has been the world's top Go player for over a decade.

And now in 2020, OpenAI released GPT-3 which is the most powerful language model ever. It can write creative fiction, generate functioning code, compose thoughtful business memos and much more. Its possible use cases are limited only by our imaginations.

Features of Machine Learning

Automation: Nowadays in your Gmail account, there is a spam folder that contains all the spam emails. You might be wondering how does Gmail know that all these emails are spam? This is the work of Machine Learning. It recognises the spam emails and thus, it is easy to automate this process. The ability to automate repetitive tasks is one of the biggest characteristics of machine learning. A huge number of organizations are already using machine learning-powered paperwork and email automation. In the financial sector, for example, a huge number of repetitive, data-heavy and predictable tasks are needed to be performed. Because of this, this sector uses different types of machine learning solutions to a great extent.

Improved customer experience: For any business, one of the most crucial ways to drive engagement, promote brand loyalty and establish long-lasting customer relationships is by providing a customized experience and providing better services. Machine Learning helps us to achieve both of them. Have you ever noticed that whenever you open any shopping site or see any ads on the internet, they are mostly about something that you recently searched for? This is because machine learning has enabled us to make amazing recommendation systems that are accurate. They help us customize the user experience. Now coming to the service, most of the companies nowadays have a chatting bot with them that are available 24x7. An example of this is Eva from AirAsia airlines. These bots provide intelligent answers and sometimes you might even not notice that you are having a conversation with a bot. These bots use Machine Learning which helps them to provide a good user experience.

Automated data visualization: In the past, we have seen a huge amount of data being generated by companies and individuals. Take an example of companies like Google, Twitter, Facebook. How much data are they generating per day? We can use this data and visualize the notable relationships, thus giving businesses the ability to make better decisions that can actually benefit both companies as well as customers. With the help of user-friendly automated data visualization platforms such as AutoViz, businesses can obtain a wealth of new insights in an effort to increase productivity in their processes.

Business intelligence: Machine learning characteristics, when merged with big data analytics can help companies to find solutions to the problems that can help the businesses to grow and generate more profit. From retail to financial services to

healthcare, and many more, ML has already become one of the most effective technologies to boost business operations.

Python provides flexibility in choosing between object-oriented programming or scripting. There is also no need to recompile the code; developers can implement any changes and instantly see the results. You can use Python along with other languages to achieve the desired functionality and results.

Python is a versatile programming language and can run on any platform including Windows, MacOS, Linux, Unix, and others. While migrating from one platform to another, the code needs some minor adaptations and changes, and it is ready to work on the new platform. To build strong foundation and cover basic concepts you can enroll in a [python machine learning](#) course that will help you power ahead your career.

Types of Machine Learning

Machine learning has been broadly categorized into three categories

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning

What is Supervised Learning?

Let us start with an easy example, say you are teaching a kid to differentiate dogs from cats. How would you do it?

You may show him/her a dog and say "here is a dog" and when you encounter a cat you would point it out as a cat. When you show the kid enough dogs and cats, he may learn to differentiate between them. If he is trained well, he may be able to recognise different breeds of dogs which he hasn't even seen.

Similarly, in Supervised Learning, we have two sets of variables. One is called the target variable, or labels (the variable we want to predict) and features (variables that help us to predict target variables). We show the program (model) the features and the label associated with these features and then the program is able to find the underlying pattern in the data. Take this example of the dataset where we want to predict the price of the house given its size. The price which is a target variable depends upon the size which is a feature.

Number of rooms	Price
1	\$100
3	\$300
5	\$500

In a real dataset, we will have a lot more rows and more than one features like size, location, number of floors and many more.

Thus, we can say that the supervised learning model has a set of input variables (x), and an output variable (y). An algorithm identifies the mapping function between the input and output variables. The relationship is $y = f(x)$.

The learning is monitored or supervised in the sense that we already know the output and the algorithm are corrected each time to optimise its results. The algorithm is trained over the data set and amended until it achieves an acceptable level of performance.

We can group the supervised learning problems as:

Regression problems – Used to predict future values and the model is trained with the historical data. E.g., Predicting the future price of a house.

Classification problems – Various labels train the algorithm to identify items within a specific category. E.g., Dog or cat (as mentioned in the above example), Apple or an orange, Beer or wine or water.

What is Unsupervised Learning?

This approach is the one where we have no target variables, and we have only the input variable(features) at hand. The algorithm learns by itself and discovers an impressive structure in the data.

The goal is to decipher the underlying distribution in the data to gain more knowledge about the data.

We can group the unsupervised learning problems as:

Clustering: This means bundling the input variables with the same characteristics together. E.g., grouping users based on search history

Association: Here, we discover the rules that govern meaningful associations among the data set. E.g., People who watch 'X' will also watch 'Y'.

What is Reinforcement Learning?

In this approach, machine learning models are trained to make a series of decisions based on the rewards and feedback they receive for their actions. The machine learns to achieve a goal in complex and uncertain situations and is rewarded each time it achieves it during the learning period.

Reinforcement learning is different from supervised learning in the sense that there is no answer available, so the reinforcement agent decides the steps to perform a task. The machine learns from its own experiences when there is no training data set present.

In this tutorial, we are going to mainly focus on Supervised Learning and Unsupervised learning as these are quite easy to understand and implement.

Machine learning Algorithms

This may be the most time-consuming and difficult process in your journey of Machine Learning. There are many algorithms in Machine Learning and you don't need to know them all in order to get started. But I would suggest, once you start practising Machine Learning, start learning about the most popular algorithms out there such as:

- Linear Regression
- Logistic Regression
- Decision Tree
- SVM
- Naive Bayes
- K-nearest neighbour
- K-Means
- Random Forest
- Gradient Boosting algorithms
 - GBM
 - XGBoost
 - LightGBM
 - CatBoost

Here, I am going to give a brief overview of one of the simplest algorithms in Machine learning, the K-nearest neighbour Algorithm (which is a Supervised learning algorithm) and show how we can use it for Regression as well as for classification. I would highly recommend checking the Linear Regression and Logistic Regression as we are going to implement them and compare the results with KNN(K-nearest neighbour) algorithm in the implementation part.

You may want to note that there are usually separate algorithms for regression problems and classification problems. But by modifying an algorithm, we can use it for both classifications as well as regression as you will see below

K-Nearest Neighbour Algorithm

KNN belongs to a group of lazy learners. As opposed to eager learners such as logistic regression, SVM, neural nets, lazy learners just store the training data in memory. During the training phase, KNN arranges the data (sort of indexing process) in order to find the closest neighbours efficiently during the inference phase. Otherwise, it would have to compare each new case during inference with the whole dataset making it quite inefficient.

So if you are wondering what is a training phase, eager learners and lazy learners, for now just remember that training phase is when an algorithm learns from the data provided to it. For example, if you have gone through the Linear Regression algorithm

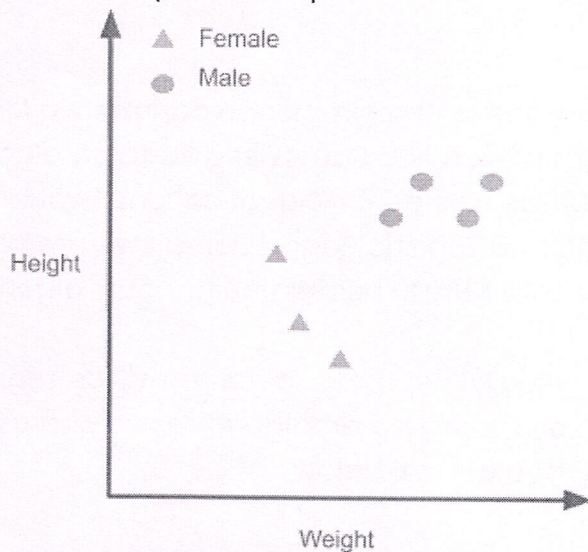
linked above, during the training phase the algorithm tries to find the best fit line which is a process that includes a lot of computations and hence takes a lot of time and this type of algorithm is called eager learners. On the other hand, lazy learners are just like KNN which do not involve many computations and hence train faster.

K-NN for Classification Problem

Now let us see how we can use K-NN for classification. Here a hypothetical dataset which tries to predict if a person is male or female (labels) on the base of the height and weight (features).

Height(cm) -feature	Weight(kg) -feature.	Gender(label)
187	80	Male
165	50	Female
199	99	Male
145	70	Female
180	87	Male
178	65	Female
187	60	Male

Now let us plot these points:

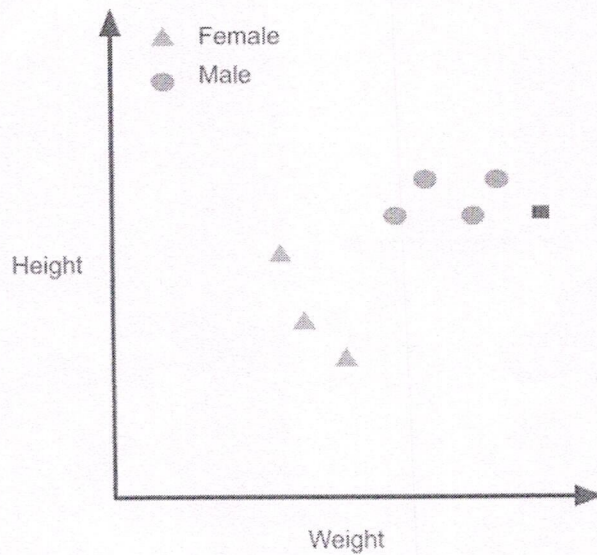


Now we have a new point that we want to classify, given that its height is 190 cm and weight is 100 Kg. Here is how K-NN will classify this point:

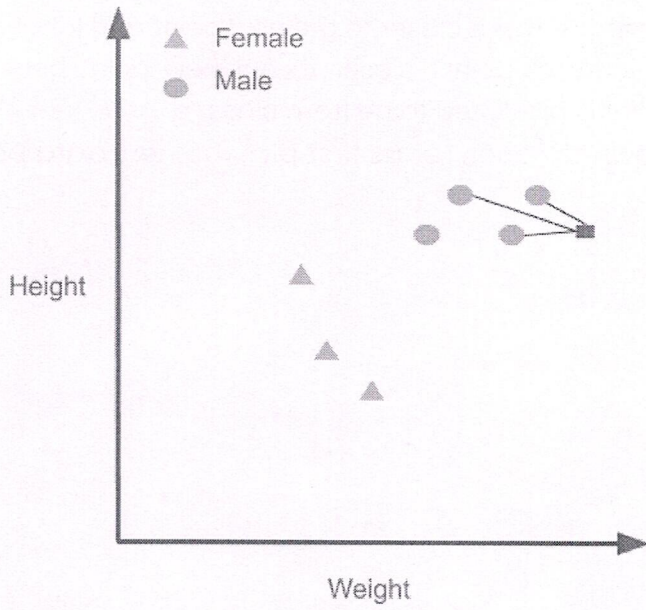
1. Select the value of K, which the user selects which he thinks will be best after analysing the data.
2. Measure the distance of new points from its nearest K number of points. There are various methods for calculating this distance, of which the most commonly known methods are - Euclidian, Manhattan (for continuous data points i.e regression problems) and Hamming distance (for categorical i.e for classification problems).

3. Identify the class of the points that are more closer to the new point and label the new point accordingly. So if the majority of points closer to our new point belong to a certain "a" class than our new point is predicted to be from class "a".

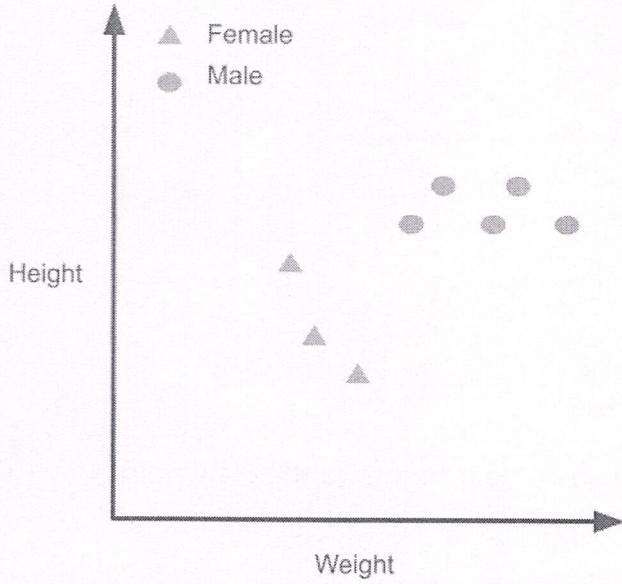
Now let us apply this algorithm to our own dataset. Let us first plot the new data point.



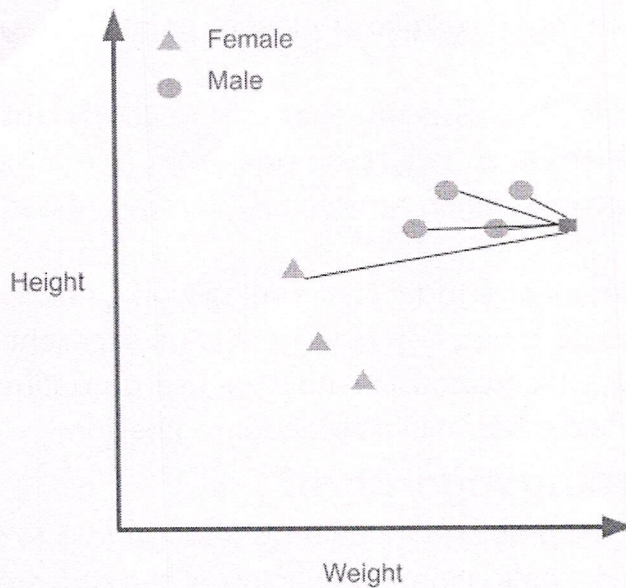
Now let us take $k=3$ i.e, we will see the three closest points to the new point:



Therefore, it is classified as Male:



Now let us take the value of $k=5$ and see what happens:



As we can see four of the points closest to our new data point are males and just one point is female, so we go with the majority and classify it as Male again. You must always select the value of K as an odd number when doing classification.

K-NN for a Regression problem

We have seen how we can use K-NN for classification. Now, let us see what changes are made to use it for regression. The algorithm is almost the same there is just one difference. In Classification, we checked for the majority of all nearest points. Here, we are going to take the average of all the nearest points and take that as predicted value. Let us again take the same example but here we have to predict the weight(label) of a person given his height(features).

Height(cm) -feature	Weight(kg) -label
187	80
165	50
199	99
145	70
180	87
178	65
187	60

Now we have new data point with a height of 160cm, we will predict its weight by taking the values of K as 1,2 and 4.

When K=1: The closest point to 160cm in our data is 165cm which has a weight of 50, so we conclude that the predicted weight is 50 itself.

When K=2: The two closest points are 165 and 145 which have weights equal to 50 and 70 respectively. Taking average we say that the predicted weight is $(50+70)/2=60$.

When $K=4$: Repeating the same process, now we take 4 closest points instead and hence we get 70.6 as predicted weight.

You might be thinking that this is really simple and there is nothing so special about Machine learning, it is just basic Mathematics. But remember this is the simplest algorithm and you will see much more complex algorithms once you move ahead in this journey.

At this stage, you must have a vague idea of how machine learning works, don't worry if you are still confused. Also if you want to go a bit deep now, here is an excellent article – Gradient Descent in Machine Learning, which discusses how we use an optimization technique called as gradient descent to find a best-fit line in linear regression.

How To Choose Machine Learning Algorithm?

There are plenty of machine learning algorithms and it could be a tough task to decide which algorithm to choose for a specific application. The choice of the algorithm will depend on the objective of the problem you are trying to solve.

Let us take an example of a task to predict the type of fruit among three varieties, i.e., apple, banana, and orange. The predictions are based on the colour of the fruit. The picture depicts the results of ten different algorithms. The picture on the top left is the dataset. The data is classified into three categories: red, light blue and dark blue. There are some groupings. For instance, from the second image, everything in the upper left belongs to the red category, in the middle part, there is a mixture of uncertainty and light blue while the bottom corresponds to the dark category. The other images show different algorithms and how they try to classified the data.

Steps in Machine Learning

I wish Machine learning was just applying algorithms on your data and get the predicted values but it is not that simple. There are several steps in Machine Learning which are must for each project.

1. **Gathering Data:** This is perhaps the most important and time-consuming process. In this step, we need to collect data that can help us to solve our problem. For example, if you want to predict the prices of the houses, we need an appropriate dataset that contains all the information about past house sales and then form a tabular structure. We are going to solve a similar problem in the implementation part.
2. **Preparing that data:** Once we have the data, we need to bring it in proper format and preprocess it. There are various steps involved in pre-processing such as data cleaning, for example, if your dataset has some empty values or abnormal values (e.g, a string instead of a number) how are you going to deal with it? There are various ways in which we can but one simple way is to just drop the rows that have empty values. Also sometimes in the dataset, we might have columns that have no impact on our

results such as id's, we remove those columns as well. We usually use Data Visualization to visualise our data through graphs and diagrams and after analyzing the graphs, we decide which features are important. Data preprocessing is a vast topic and I would suggest checking out [this article](#) to know more about it.

3. **Choosing a model:** Now our data is ready is to be fed into a Machine Learning algorithm. In case you are wondering what is a Model? Often "machine learning algorithm" is used interchangeably with "machine learning model." A model is the output of a machine learning algorithm run on data. In simple terms when we implement the algorithm on all our data, we get an output which contains all the rules, numbers, and any other algorithm-specific data structures required to make predictions. For example, after implementing Linear Regression on our data we get an equation of the best fit line and this equation is termed as a model. The next step is usually training the model in case we don't want to tune hyperparameters and select the default ones.
4. **Hyperparameter Tuning:** Hyperparameters are crucial as they control the overall behaviour of a machine learning model. The ultimate goal is to find an optimal combination of hyperparameters that gives us the best results. But what are these hyper-parameters? Remember the variable K in our K-NN algorithm. We got different results when we set different values of K. The best value for K is not predefined and is different for different datasets. There is no method to know the best value for K, but you can try different values and check for which value do we get the best results. Here K is a hyperparameter and each algorithm has its own hyperparameters and we need to tune their values to get the best results. To get more information about it, check out [this article -Hyperparameter Tuning Explained](#).
5. **Evaluation:** You may be wondering, how can you know if the model is performing good or bad. What better way than testing the model on some data. This data is known as testing data and it must not be a subset of the data (training data) on which we trained the algorithm. The objective of training the model is not for it to learn all the values in the training dataset but to identify the underlying pattern in data and based on that make predictions on data it has never seen before. There are various evaluation methods such as K-fold cross-validation and many more. We are going to discuss this step in detail in the coming section.
6. **Prediction:** Now that our model has performed well on the testing set as well, we can use it in real-world and hope it is going to perform well on real-world data.

Evaluation of Machine learning Model

For evaluating the model, we hold out a portion of data called test data and do not use this data to train the model. Later, we use test data to evaluate various metrics.

The results of predictive models can be viewed in various forms such as by using confusion matrix, root-mean-squared error (RMSE), AUC-ROC etc.

A confusion matrix used in classification problems is a table that displays the number of instances that are correctly and incorrectly classified in terms of each category within the attribute that is the target class as shown in the figure below:

TP (True Positive) is the number of values predicted to be positive by the algorithm and was actually positive in the dataset. TN represents the number of values that are expected to not belong to the positive class and actually do not belong to it. FP depicts the number of instances misclassified as belonging to the positive class thus is actually part of the negative class. FN shows the number of instances classified as the negative class but should belong to the positive class.

Now in Regression problem, we usually use RMSE as evaluation metrics. In this evaluation technique, we use the error term.

Let's say you feed a model some input X and the model predicts 10, but the actual value is 5. This difference between your prediction (10) and the actual observation (5) is the error term: $(f_{\text{prediction}} - i_{\text{actual}})$. The formula to calculate RMSE is given by:

Where N is a total number of samples for which we are calculating RMSE.

In a good model, the RMSE should be as low as possible and there should not be much difference between RMSE calculated over training data and RMSE calculated over the testing set.

Python for Machine Learning?

Although there are many languages that can be used for machine learning, according to me, Python is hands down the best programming language for Machine Learning applications. This is due to the various benefits mentioned in the section below. Other programming languages that could be used for Machine Learning Applications are R, C++, JavaScript, Java, C#, Julia, Shell, TypeScript, and Scala. R is also a really good language to get started with machine learning.

Python is famous for its readability and relatively lower complexity as compared to other programming languages. Machine Learning applications involve complex concepts like calculus and linear algebra which take a lot of effort and time to implement. Python helps in reducing this burden with quick implementation for the Machine Learning engineer to validate an idea. You can check out the [Python Tutorial](#) to get a basic understanding of the language. Another benefit of using Python in Machine Learning is the pre-built libraries. There are different packages for a different type of applications, as mentioned below:

1. Numpy, OpenCV, and Scikit are used when working with images
2. NLTK along with Numpy and Scikit again when working with text
3. Librosa for audio applications

4. Matplotlib, Seaborn, and Scikit for data representation
5. TensorFlow and Pytorch for Deep Learning applications
6. Scipy for Scientific Computing
7. Django for integrating web applications
8. Pandas for high-level data structures and analysis

Implementation of algorithms in Machine Learning with Python

Before moving on to the implementation of machine learning with Python part, you need to download some important software and libraries. For this, you can refer to [this article](#) where you will know how to install anaconda on your PC and basics of Python. Anaconda is an open-source distribution that makes it easy to perform Python/R data science and machine learning on a single machine. It contains all most all the libraries that are needed by us. In this tutorial, we are mostly going to use the `scikit-learn` library which is a free software machine learning library for the Python programming language. Now, we are going to implement all that we learnt till now. We will solve a Regression problem and then a Classification problem using the seven steps mentioned above.

Implementation of a Regression problem

We have a problem of predicting the prices of the house given some features such as size, number of rooms and many more. So let us get started:

1. **Gathering data:** We don't need to manually collect the data for past sales of houses. Luckily there are some good people who do it for us and make these datasets available for us to use. Also let me mention not all datasets are free but for you to practice, you will find most of the datasets free to use on the internet.

The dataset we are using is called the Boston Housing dataset. Each record in the database describes a Boston suburb or town. The data was drawn from the Boston Standard Metropolitan Statistical Area (SMSA) in 1970. The attributes are defined as follows (taken from the UCI Machine Learning Repository).

1. CRIM: per capita crime rate by town
2. ZN: proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS: proportion of non-retail business acres per town
4. CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
5. NOX: nitric oxides concentration (parts per 10 million)
6. RM: average number of rooms per dwelling
7. AGE: the proportion of owner-occupied units built prior to 1940
8. DIS: weighted distances to five Boston employment centers

9. RAD: index of accessibility to radial highways
10. TAX: full-value property-tax rate per \$10,000
11. PTRATIO: pupil-teacher ratio by town
12. B: $1000(Bk-0.63)^2$ where Bk is the proportion of blacks by town
13. LSTAT: % lower status of the population
14. MEDV: Median value of owner-occupied homes in \$1000s

Here is a link to download this dataset.

Now after opening the file you can see the data about House sales. This dataset is not in a proper tabular form, in fact, there are no column names and each value is separated by spaces. We are going to use Pandas to put it in proper tabular form. We will provide it with a list containing column names and also use delimiter as '\s+' which means that after encountering a single or multiple spaces, it can differentiate every single entry.

We are going to import all the necessary libraries such as Pandas and NumPy. Next, we will import the data file which is in CSV format into a pandas DataFrame.

```

1 import numpy as np
2 import pandas as pd
3 column_names = ['CRIM', 'ZN', 'INDUS', 'CHAS', 'NOX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX',
4 bos1 = pd.read_csv('housing.csv', delimiter=r"\s+", names=column_names)

```

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	MEDV
0	0.00632	18.0	2.31	0	0.538	6.575	65.2	4.0900	1	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0	0.469	6.421	78.9	4.9671	2	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0	0.469	7.185	61.1	4.9671	2	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0	0.458	6.998	45.8	6.0622	3	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0	0.458	7.147	54.2	6.0622	3	222.0	18.7	396.90	5.33	36.2
5	0.02985	0.0	2.18	0	0.458	6.430	58.7	6.0622	3	222.0	18.7	394.12	5.21	28.7
6	0.08829	12.5	7.87	0	0.524	6.012	66.6	5.5605	5	311.0	15.2	395.60	12.43	22.9
7	0.14455	12.5	7.87	0	0.524	6.172	96.1	5.9505	5	311.0	15.2	396.90	19.15	27.1
8	0.21124	12.5	7.87	0	0.524	5.631	100.0	6.0821	5	311.0	15.2	386.63	29.93	16.5

2. Preprocess Data: The next step is to pre-process the data. Now for this dataset, we can see that there are no NaN (missing) values and also all the data is in numbers rather than strings so we won't face any errors when training the model. So let us just divide our data into training data and testing data such that 70% of data is training data and the rest is testing data. We could also scale our data to make the predictions much accurate but for now, let us keep it simple.

3. Choose a Model: For this particular problem, we are going to use two algorithms of supervised learning that can solve regression problems and later compare their results. One algorithm is K-NN (K-nearest Neighbor) which is explained above and the other is [Linear Regression](#). I would highly recommend to check it out in case you haven't already.

4. Hyperparameter Tuning: Since this is a beginners tutorial, here, I am only going to turn the value of K in the K-NN model. I will just use a for loop and check results of k ranging from 1 to 50. K-NN is extremely fast on small dataset like ours so it won't take any time. There are much more advanced methods of doing this which you can find linked in the steps of Machine Learning section above.

5. Evaluating the model: For evaluating the model we are going to use the `mean_squared_error()` method from the scikit-learn library. Remember to set the parameter 'squared' as False, to get the RMSE error.

6. Prediction: Now we can use the models to predict the prices of the houses using the predict function as we did above. Make sure when predicting the prices that we are given all the features that were present when training the model.

Advantages of Machine Learning

1. Easily identifies trends and patterns

Machine Learning can review large volumes of data and discover specific trends and patterns that would not be apparent to humans. For instance, for e-commerce websites like Amazon and Flipkart, it serves to understand the browsing behaviors and purchase histories of its users to help cater to the right products, deals, and reminders relevant to them. It uses the results to reveal relevant advertisements to them.

2. Continuous Improvement

We are continuously generating new data and when we provide this data to the Machine Learning model which helps it to upgrade with time and increase its performance and accuracy. We can say it is like gaining experience as they keep improving in accuracy and efficiency. This lets them make better decisions.

3. Handling multidimensional and multi-variety data

Machine Learning algorithms are good at handling data that are multidimensional and multi-variety, and they can do this in dynamic or uncertain environments.

4. Wide Applications

You could be an e-tailer or a healthcare provider and make Machine Learning work for you. Where it does apply, it holds the capability to help deliver a much more personal experience to customers while also targeting the right customers.

Disadvantages of Machine Learning

1. Data Acquisition

Machine Learning requires a massive amount of data sets to train on, and these should be inclusive/unbiased, and of good quality. There can also be times where we must wait for new data to be generated.

2. Time and Resources

Machine Learning needs enough time to let the algorithms learn and develop enough to fulfill their purpose with a considerable amount of accuracy and relevancy. It also needs massive resources to function. This can mean additional requirements of computer power for you.

3. Interpretation of Results

Another major challenge is the ability to accurately interpret results generated by the algorithms. You must also carefully choose the algorithms for your purpose. Sometimes, based on some analysis you might select an algorithm but it is not necessary that this model is best for the problem.

4. High error-susceptibility

Machine Learning is autonomous but highly susceptible to errors. Suppose you train an algorithm with data sets small enough to not be inclusive. You end up with biased predictions coming from a biased training set. This leads to irrelevant advertisements being displayed to customers. In the case of Machine Learning, such blunders can set off a chain of errors that can go undetected for long periods of time. And when they do get noticed, it takes quite some time to recognize the source of the issue, and even longer to correct it.

Future of Machine Learning

Machine Learning can be a competitive advantage to any company, be it a top MNC or a startup. As things that are currently being done manually will be done tomorrow by machines. With the introduction of projects such as self-driving cars, Sophia (a humanoid robot developed by Hong Kong-based company Hanson Robotics) we have already started a glimpse of what the future can be. The Machine Learning revolution will stay with us for long and so will be the future of Machine Learning.