Leading University

Department of Computer Science and Engineering

Course Title:	Machine Learning
Course Code:	CSE-4233
Assignment Title:	Assignment on Midterm Preparatory Assignmnet

Submitted To:

Name:	Shafkat Kibria
Designation:	Assistant Professor
Faculty:	Dept. of CSE

Submitted By:

Name:	Md. Asadullah-Al-Galib
Department:	Dept. of CSE
Batch:	53rd
Section:	G
ID:	2012020303

Submission Date: 2023-03-18

1) MISP refers Million instruction per second.

2) In human brain there is 10^{11} nurons and 10^{12-13} synapse.

3) Machine Learning: It is a field of study that gives the ability to computers to learn without being explicitly programed.

we need machine learning because it gives handling large amount of datas, Automative repetitive tasks, improving decision making, personalization, Scaling UP.

There are certain functions which cannot be programed as they are not relavent mathmatical equations describing the function. ex. The reason why human cliess add.

4) Some machine learning methods are:-
Artificial neural networks, Fuzzy logic, Genetic Algorithm, Support vector machine, Decision tree, Bayesian networks

Artificial neural networks: This models inspired by the structure of the neural system.

Fuzzy logic: Fuzzy set theory, which can thought of as a generalization of classical set theory.

Genetic Algorithm: The primary purpose of using this algorithm is optimization through evaluation on fitness function.

Support vector machine: It is designed from Statical learning theory.

Bayesian networks: It's use for reasoning under uncertainty. It is used for ML task such as classifications, regression. Bayesian network is trained as.

5) There are some examples of ML base application in current context –

Applications can't program by hand ex. Autonomous helicopter, handwriting recognization, Natural language processing.

Handwriting recognization. There are some numbers that human maybe recognize is it number or anything but ML can recognize the number easily.

Autonomus helicopter it will fly without human touch.

Understanding human learning by brain and Reed AI. Self customization programs ex: Amazon, Netflex recomended movies or products.

6) Supervised learning is a type of machine learning where an Algorithm trained ~~with~~ on a labled dataset. which means desired output is already known. The goal of supervised learning is to learn a mapping between inputs data or features and the output label by finding patterns and relationship in the training data. example.

6. Pizza sales based on inch (size) and with Price.

Size	Price
8" inch	1000+
12 inch"	1200+.

here size is features and Price is label. here Price is the known output. It's a supervised learning.

7) Unsupervised learning is a type of machine learning where the algorithm trained on a features dataset. which means the desired data set is unknown to the algorithm. It can be only predict on a desired input. ~~example — Pizza size~~ It is a clustering problem. There is no grouping of dataset. example: My dataset has many animals characteristic. dog, cat, cow. now. unsupervise learning identify the all characteristics and group them. when I input any characteristic it will answer ~~the~~ what is it. dog, cat or cow.

8) The hypotheses in machine learning ~~com~~. that maps input features to output label. The goal of machine learning algorithm is to learn ~~bg~~ hypothesis function that can actuals predict output label for new, unseen ~~lab~~ input data.

9) Univariate linear regression is a type of linear regression that involves a single input features and single output label. The goal the learn a linear relationship between the input features and output label.

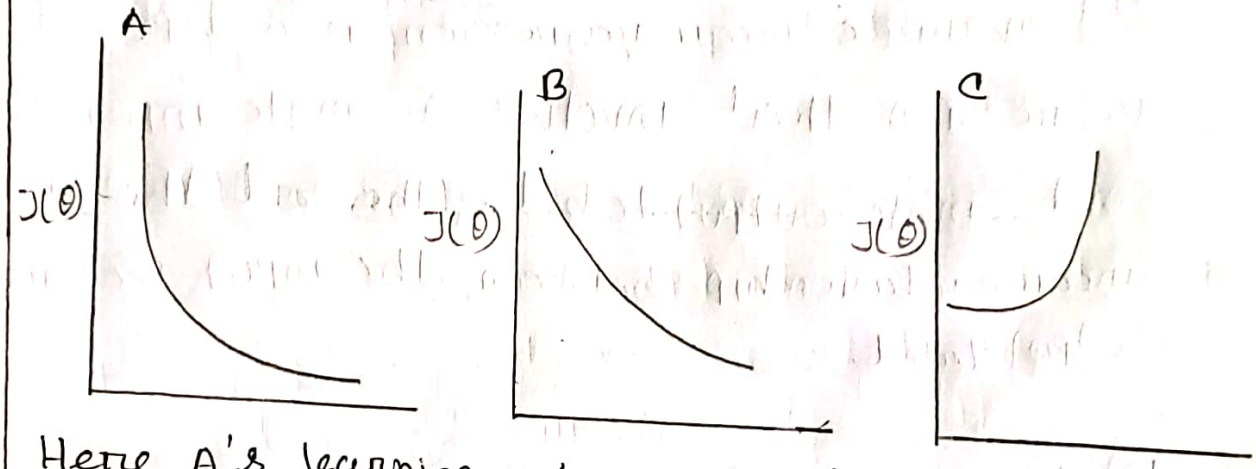
10) In machine learning a cost function is a measure of the difference between the predicate output and the actuals output label in the training data.

cost function $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} h_\theta((x^{(i)}) - y^{(i)})^2$

the goal is minimize the cost function $J(\theta_0, \theta_1)$.

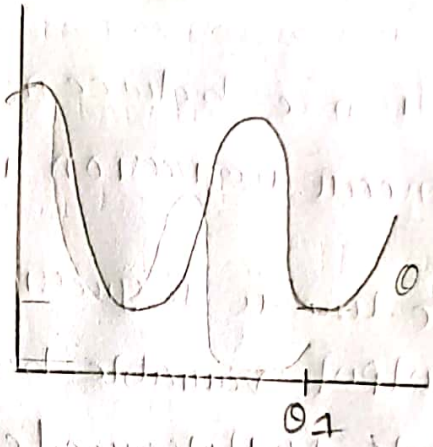
11) The name of the algorithm is used to minimize the cost function is Gradient descent.

12) The effect of learning rate α in Gradient decent if is to small it's very slow converge and if α is to large it's may not decrease in every iteration.



Here A's learning α is 0.1, B's 0.01 and C's 1. we can see here A's learning rate is Perfect. it's converge to global minima, here B's learning rate is too small, that's why it's slowly converge and C's is 1. It's may not decrese in every iteration. It maybe converge or diverge but not come in global minima

13) If gradient descent converge to local or global minimum we will not update θ. θ will keep unchanged, because if we drew a slop here the value of slope is 0. We know the goal of Gradient descent is decrease the cost function. In local minimum it's not to be increase, so that we will keep unchanged the θ.



here θ_1 is in the local minimum. and here the slop is 0. and our goal is minimize the cost. so θ_1 will keep unchanged.

14) We should keep learning rate α fixed. because if α value is too small it's slowly decrease and if α value is too large it's may not go to the global minimum. If we change learning rate every iteration the cost may not decrease or decrease very slowly so we should keep learning rate α fixed.

15) The difference between Univariate and Multivariate linear regression is-

(i) Univariate linear regression involves predicting a single output variable based on one ~~out~~ input variable, while Multivariate linear regression involves predicting the output variable based on multiple input variables.

(ii) In Univariate linear regression there is only one independent input variable, while in multivariate there is multiple input variable.

(iii) Univariate linear regression can only model linear relationship between variables. while ~~no~~ Multivariate linear regression can model non-linear ~~regression~~ relationship between variables.

(iv) Multivariate linear regression allows for better prediction accuracy when there are multiple factors influencing the outcome variables.

example: 1 first example for univariate and 2nd one for Multivariate linear regression.

Size of House (Feet)	Price(BDT)
2014	460000
1416	232000
1534	315000
852	178000

fig: Univariate

Ram (GB)	ROM (GB)	Camera (MP)	Price (BDT)
2	160	8	10.000
4	32	8	12.000
68	64	12	19.000
88	128	48	35.000

fig: Multivariate

16) Features scaling is a technique used in machine learning to standerized the range of independent variables or features of dataset, which helps in improving the performance and convergence of many machine learning algorithm

we need ~~need~~ Feature scaling in machine learning many machine learning algorithms are sensitive to the scale of features and features scaling helps avoiding dominance of certains features over others. Features scaling also helps in reducing the computational resources required for training the model and improving.

17) In one variable linear regression, feature scaling not necessary since there is only one independent variable. ~~However~~ However. if the range of values for the independent variable is large, it might be useful to scale the variable to improve the performance of the algorithm. Feature scaling is most commonly used for multivariate linear regression.

18) After applying features scaling, the range of every feature should ideally be between -1 and 1 or between 0 and 1. However the exact range will depend on the type of scaling technique used (such as min-max scaling). The most important fact is to ensure that the range of all features is consistence and on a similar scale to avoid any features dominating the model.

19) Mean normalization –

$$x_1 = \frac{x_1 - x'}{s_1}$$

$$= \frac{\text{age of the house} - 38}{(50 - 30)}$$

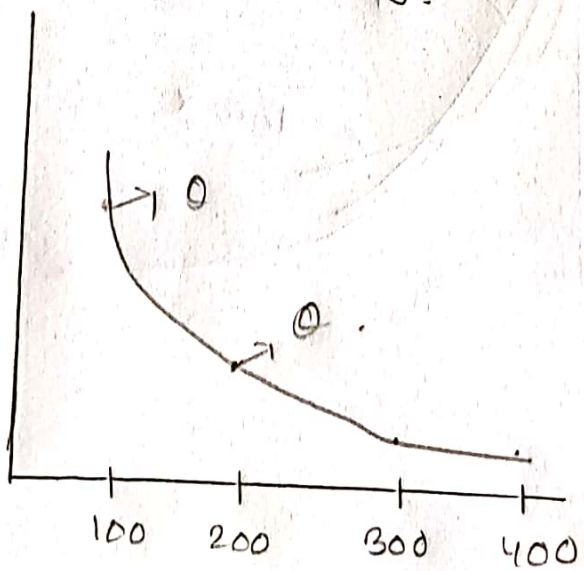
$$= \frac{\text{age of the house} - 38}{20}$$

x' = average value of x_1 in training.

s_1 = range (min-max)

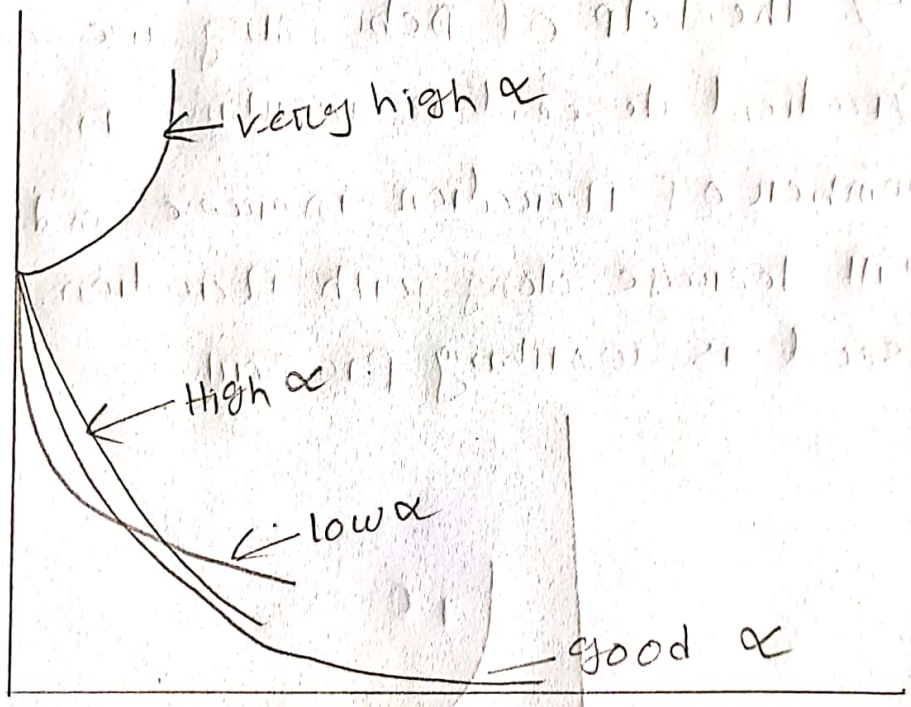
x_1 = value / features.

20) The help of Debugging we can decide Gradient descent is working Properly If number of iteration increase and $J(\theta)$'s value will decrease along with iteration. So our gradiant descent is working properly.



21) We know if gradient descent ~~descent~~ works properly that cost function should decrease after each iteration. But when gradient descent can no longer reduce the cost function and remains at more less the same level, it has converged and it may sometimes require a large change in the number of iterations to converge.

If the plots shows that learning curve just goes up and down without reaching a lower point then we reduce the learning rate. Also try 0.001, 0.003 0.01, 0.03, 0.1, 0.3, 1, as the learning rate and see which one performs best. If than the gradient descent not working properly then we need to random restart



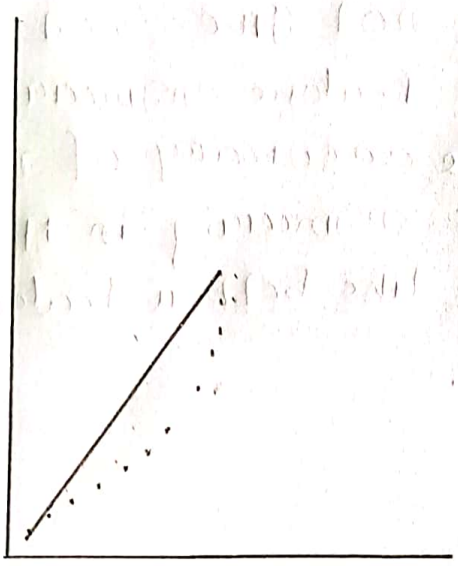
22) we will automatically choose the learning rate α. to be (0.001, 0.003, 0.01, 0.03, 0.001, 0.3, 1) and plot the $J(\theta)$ against the number of iterations for each to see which value of α gives the best converge.

23) Feature Engineering: It is a pre-process step of machine learning. which extracts features from raw data.

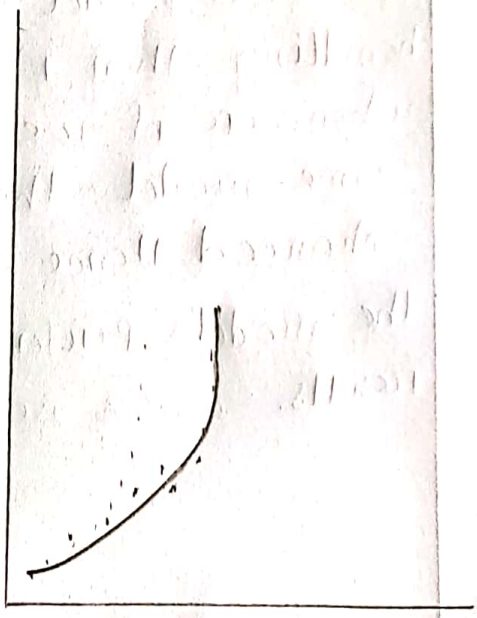
In machine learning, the performance of the model depends on data pre-processing and data handling, But if we create a model without pre-processing or data handling, than it may not give good accuracy. Whenever if we apply feature engineering on the same model, then the accuracy of model is enhanced, Hence feature engineering in ML improves the model's performance like better features better results. So we need it.

24) Polynomial Regression: It is a form of ~~linear~~ regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth layer polynomial in x.

If we apply a linear model on a linear dataset, then it provides us a good result as we have seen in simple linear regression. But if we apply the same model without any modifications on a non-linear dataset, then it will produce a drastic output. Due to which loss function increase, the occure rate will be high and accuracy will be decrease. so for such cases datapoints arranged in a non linear ~~fun~~. we need polynomial regression model.



Simple linear model



Polynomial model.

25) Logistic regression is a supervised learning that use to calculate or predict the probability of a binary (0/1) event occuring.

sometimes our data contains outlins in which case linear regression fails to produce a line of best fit based on the data points. Logistic ~~function~~ regression used sigmoed function and gives output in binary.

example: this numbers is there or five. a simple linear can't identify it but logistic function identify it and used sigmoed function for giving binary value. there is a decision boundary <50% is 0 and >50 is 1.

26) Sigmoed function is used for the logistic regressio

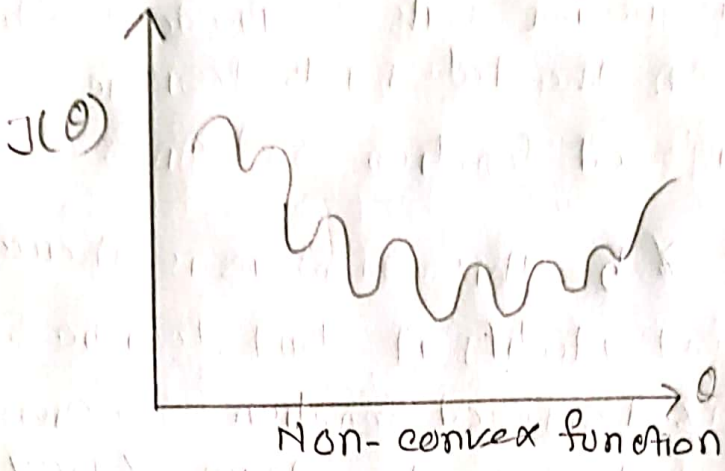
$$h_\theta(x) = \frac{1}{1 + e^{-z}} \quad ; 0 \leq h_\theta(x) \leq 1.$$

27) Decision boundary is a surface that separates data points belonging different class lables.

28) Hypothesis function for non-linear decisions boundaries.

$$h_\theta(x) = g(\theta_0 + \theta_1 x + \theta_2 x_2 + \theta_3 x_3^3 + \theta_4 x_4^4 \ldots)$$

29) If we try to use the cost function of the linear regression in "Logistic Regression" then it will be end up as a non-convex function with many local minima, where it will be very difficult to minimize the cost value and find the global minimum.



Non-convex function

So, Logistic regression, the cost function is defined as

$$\text{cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & ; \text{if } y = 1. \\ -\log(1 - h_\theta(x)) & ; \text{if } y = 0 \end{cases}$$

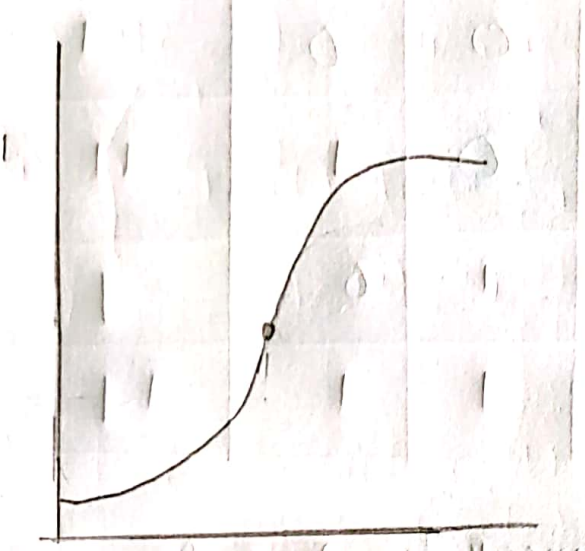
30) The shape of the "cost function" for logistic regression is "S" shape.

we use sigomed function in logistic regression, here is a decision boundary it's give us bineiry answer.

$h_\theta(x) = g(\theta^T x)$

$$g(z) = \frac{1}{1+e^{-z}}$$

if $g(z)$ is $>50\%$ it's gives as '1' and if $y(z)$ is $<50\%$ it's gives 0.



31) The activations functions used in simple perceptron is - Treashold step, sign and sigmoed functions.

we know that,

$$\text{Total input, } Si = \sum_{i=1}^{n} i_j n w_n + b$$

And Treashold activation function,

$$x_j = f(s_j) = \begin{cases} 1 ; s_j \geq 0 \\ 0 ; s_j < 0 \end{cases}$$

The OR function truth table -

x_1	x_2	f
0	0	-1
0	1	+1
1	0	-1
1	1	+1

The linear function for OR 2 inputs.

$$W^T x = w_0 + w_1 x_1 + w_2 x_2$$

Let $W = \begin{bmatrix} -1 \\ 0.6 \\ 0.6 \end{bmatrix}$ and $\eta = 0.5$

This one is correctly classified, no need action.

Now,

When $x_1 = 0$, $x_2 = 1$

$$= -0.5 + 0.6 \times 0 + 1.1 \times 1$$
$$= 0.6$$

This one is correctly classified.

When $x_1 = 1$, $x_2 = 0$

$$= -0.5 + 0.6$$
$$= 0.1$$

This one is correctly classified.

When $x_1 = 1$, $x_2 = 1$

$$= -0.5 + 0.6 + 1.1$$
$$= 1.2$$

This one is correctly classified.

32) The types of Machine learning is Supervised learning, unsupervised learning, Reinforcement learning.

In Supervised learning the common used are classification and Regression.

In Unsupervised learning the common uses are clustering and Dimentionally Reduction.

In ~~Rein~~ Reinforcement learning the common uses are Game AI and Robotics.

33) The Steges of Machine learning process.

① Data collection:- The first stege of machine learning project is data collecting. This involves gathering the relavent data needed to train machine learning model.

⑪ Data Preparation:- After data collected, it must preprocessed and cleened to remove any missing or irrelevent data.

(iii) Model Building (Step):- Onced data is prepared, the next is to build machine learning model.

(iv) Model evaluation:- After model is trained, it must be evaluated to determine it's performance and effectiveness.

(v) Model optimization:- After evaluating, the model need to be optimized.

(vi) Deployment: After optimized it ~~need to be~~ can deploy.

(vii) Monitoring: After the deployment it need to be monitoring, tracking performance.

34) The purpose of cost function learning is to measure how well a model fits the training data. It's provides a quantifiable measure of the diffrence between the predicted output and the actual output. The goal of the model is to minimize the cost function to improve it's accuracy on new, unseen data.

cost function $J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})$

The common cost function calculating method is "Gradient descent".