Page No.:
Date:

Machine Learning (CS3EA07)

Assignment - 1

Bhumi Gupta
EN19CS301094
V 'B'

Q.1 Define Machine Learning. Explain different perspectives and issues in machine learning.

Ans Machine learning is an application of Artificial Intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed.

Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves.

The process of learning begins with observations or data, such as direct experience, or instructions. The primary aim is to allow the computers learn automatically without human intervention or assistance and adjust actions accordingly.

Issues in Machine learning-

Machine learning provides businesses with the knowledge to make more informed, data-driven decisions that are faster than traditional approaches. Here are 5 common issues.

1) Understanding Which Processes Need Automation.

It's becoming increasingly difficult to separate fact from fiction in terms of Machine learning today. Before you decide on which AI platform to use, you need to evaluate which problems you're seeking to solve. The easiest processes to automate are the ones that are done manually every day with no variable output.

2) Lack of Quality Data.

The no. 1 of problem facing Machine Learning is the Lack of good data. While enhancing algorithms often consumes most of the time of developers in AI, data quality is essential for the algorithms to function as intended. The solution to this conundrum is to take the time to evaluate and scope data with meticulous data governance, data integration, and data exploration until you get clear data.

3) Inadequate Infrastructure.

ML requires vast amounts of data churning capabilities. Legacy systems often can't handle the workload and buckle under pressure. You should check if your infrastructure can handle ML. If it can't, you should look to upgrade, complete with hardware acceleration and flexible storage.

4) Implementation.

Organizations often have analytics engines working with them by the time they choose to upgrade to ML. Integrating newer ML methodologies into existing methodologies is a complicated task. Maintaining proper interpretation and documentation goes a long way to easing implementation. Partnering with an implementation partner can make the implementation of services like anomaly detection, predictive analysis, and ensemble modeling much easier.

5) Lack of Skilled Resources.

Deep analytics and ML is their current forms are still new technologies. Thus, there is a shortage of skilled employees available to manage and develop analytical content for ML. Data scientists often need a combination of Domain experience

as well as in-depth knowledge of science, technology, and mathematics. Recruitment will require you to pay large salaries as these employees are often in high-demand and know their worth. You can also approach your vendor for staffing help as many managed service providers keep a list of skilled data scientists to deploy anytime.

Q.2 Explains the steps involved in designing a learning system.

Ans Steps for designing learning systems are :-

Step-1 → Choosing the Training Experience :
The very important and first task is to choose the training data or training experience which will be fed to the ML algorithm.
The attributes which will impact on success and failure of data :

- The training experience will be able to provide direct or indirect feedback regarding choices.
- Second important attribute is the degree to which the learner will control the sequences of training examples.
- Third important attribute is how it will represent the distribution of examples over which performance will be measured.

Step-2 → Choosing target function :
The next important step is choosing the target function. It means according to the knowledge fed to the algorithm the machine learning will choose nextMove function which will describe what type of legal moves should be taken.

Step-3 → Choosing Representation for Target function :
When the machine algorithm will know all the possible legal moves the next step is to choose the optimized move using

any representation i.e. using linear equations, Hierarchical Graph Representation, Tabular form etc. The NextMove function will move the Target move like out of these move which will provide more success rate.

Step-4 → Choosing Function Approximation Algorithm..
An optimized move cannot be chosen just with the training data. The training data had to go through with set of example and through these examples the training data will approximates which steps are chosen and after that machine will provide feedback on it.

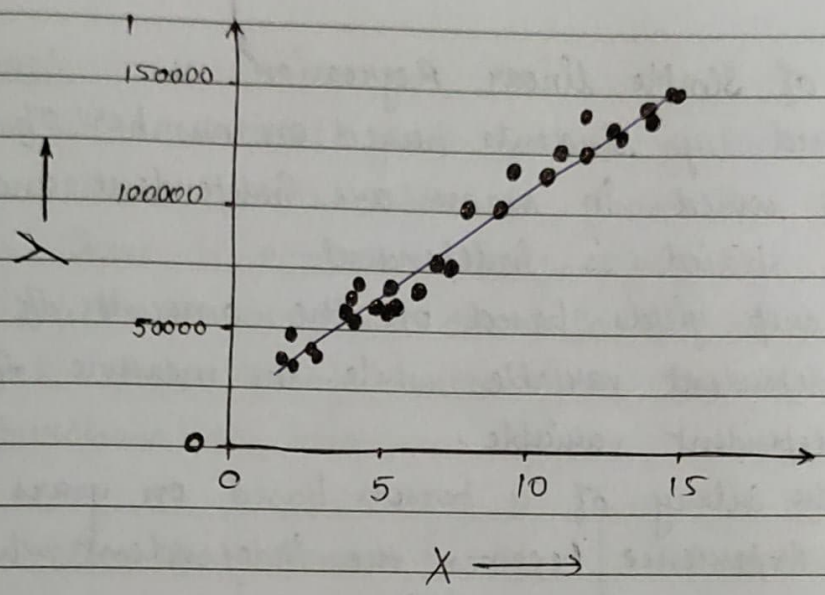
Step-5 → Final Design.
The final design is created at last when system goes from number of examples, failures and success, correct and incorrect decision and what will be the next step, etc.
Example :- DeepBlue is an intelligent computer which is ML-based won chess game against the chess expert Gravry Kasparov, and it became the first computer which had beaten a human chess expert.

Q.3 Explain linear regression and logistic regression model with example. Describe the application areas for both the models.

Ans Linear Regression Model →
Linear Regression is a machine learning algorithm based on supervised learning. It perfoms a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on - the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.

Bhumi Gupta

Page No.:
Date:



Example

It can be used to quantify the relative impacts of age, gender, and diet on height.

The table below shows some data from the early days of the Italian clothing company Benetton.

If we use advertising as the predictor variable, linear regression estimates that **Sales = 168 + 23 Advertising**.

Year	Sales	Advertising
1	651	23
2	762	26
3	856	30
4	1063	34
5	1190	43
6	1298	48
7	1421	52
8	1440	57
9	1518	58

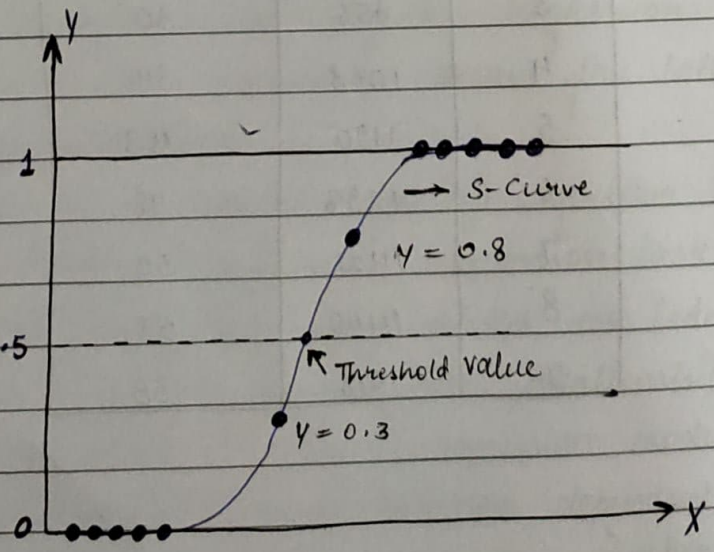
Applications of Simple Linear Regression.

1) Marks scored by students based on number of hours studied. Here marks scored in exams are independent and the number of hours studied is independent.
2) Predicting crop yields based on the amount of rainfall. Yield is a dependent variable while the measure of precipitation is an independent variable.
3) Predicting the salary of a person based on years of experience. Therefore, Experience becomes the independent while Salary turns into the dependent variable.

Logistic Regression Model →

Logistic regression is one of the most popular Machine learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

It predicts the output of a categorical dependent variable. Therefore the outcome must be a categorical or discrete value. It can be either Yes or No, 0 or 1, true or false, etc. but instead of giving the exact value as 0 and 1, it gives the probabilistic values which lies between 0 and 1.



Example

There is a dataset given which contains the information of various users obtained from the social networking sites. There is a car making company that has recently launched a new SUV car. So the company wanted to check how many users from the dataset, wants to purchase the car.

User ID	Gender	Age	EstimatedSalary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	0
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	0
15728773	Male	27	58000	0
15598044	Female	27	84000	0
15694829	Female	32	150000	1

Applications of Logistic Regression.

Logistic regression is used in various fields, including machine learning, most medical fields, and social sciences.

For Ex:- the Trauma and Injury Severity Score, which is widely used to predict mortality in injured patients, was originally developed by Boyd et al. using Logistic regression. Many other medical scales used to assess severity of a patient have been developed using logistic regression. Logistic reg. may be used to predict the risk of developing a given disease, based on observed characteristics of the patients.

Q.4 Explain how gradient decent algorithm works to optimize cost function for linear regression.

Ans Gradient Descent is the process of minimizing a function by following the gradients of the cost function.
This involves knowing the form of the cost as well as the derivatives so that from a given point you know the gradient and can move in that direction, e.g. downhill towards the minimum value.
In ML we can use a similar technique called stochastic gradient descent to minimize the error of a model on our training data.
The way this works is that each training instance is shown to the model one at a time. The model makes prediction for a training instance, the error is calculated and the model is updated in order to reduce the error for the next prediction.

This procedure can be used to find the set of coefficients in a model that result in the smallest error for the model on the training data. Each iteration the coefficients, called weights in ML language are updated using the equation:

$$W = W - \text{alpha} * \text{delta}$$

where, w is the coefficient or weight being optimized.
alpha is a learning rate
gradient is the error for the model on the training data attributed to the weight.

Q.5 Explain k-nearest Neighbor learning algorithm with example.
Ans K-Nearest Neighbour is one of the simplest ML algorithms based on supervised learning technique.
k-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
k-NN algorithm stores all the variable data and classifies a

new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K-NN algorithm.
K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data.
It is also called a lazy learner algorithm because it does not learn from the training set.

Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category.