**FSDS MAY BATCH 2022(ML Assignment -3)**

**Submitted by: Shubham Tiwari**

Q1:Explain the term machine learning, and how does it work? Explain two machine learning applications in the business world. What are some of the ethical concerns that machine learning applications could raise?

Ans: Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://www.techtarget.com/whatis/definition/algorithm) use historical data as input to predict new output values. Some of the most common case include mail-spam detection, malware threat detection and predictive maintenance.Machine learning provides statistical tool to analyze,visualize, perform, prediction and other task with the help of data.

It basically works as we know that it basically uses two kind of techniques i.e **supervised learning**  which trains a model on known input and output data so that it can predict future outputs and **unsupervised learning** which finds hidden patterns or internal structures in input data. Basic steps include as :

1. Collecting Data.
2. Preparing and Visualizing the Data.
3. Choosing a Model.
4. Training the Model.
5. Evaluating the Model.
6. Hyperparameter Tuning.
7. Making Predictions.

Two Machine learning applications in the business world:

**1) Image Recognition**

Image recognition involves machine learning through the use of data mining, computer vision, pattern recognition, and database knowledge discovery to produce numeric and symbolic information from images and other multi-dimensional data.

**2) Medical Diagnosis**

Machine learning in healthcare can potentially make near perfect diagnosis because of the objectivity in data analysis.

**Some of the ethical concerns that machine learning algorithms can raise are:**

1) Firstly it can be expensive. Machine learning projects are typically driven by data scientists, who command high salaries. These projects also require software infrastructure that can be expensive.

2) There is also the problem of machine learning bias. Algorithms trained on data sets that exclude certain populations or contain errors can lead to inaccurate models of the world that, at best, fail and, at worst, are very poor.

Q2: Describe the process of human learning:

i. Under the supervision of experts

Ans: Human learning under the supervision of experts is a type of **supervised learning**, which uses a set of human-labeled training data to develop a model.Here subject matter experts accelerate the learning process by teaching the technology in real-time. For example, if the machine learning model comes across a piece of data it is uncertain about, a human can be asked to weigh in and give feedback. The model then learns from this input, and uses it to make a more accurate prediction the next time. Expert-guided machine learning works from the bottom up by first using algorithms to conduct the heavy lifting of identifying relationships within the data, and engaging humans when necessary for training or validation. This means that, the amount of time a human needs to spend performing a specific task will decrease as the machine learning accuracy increases.

ii. With the assistance of experts in an indirect manner.

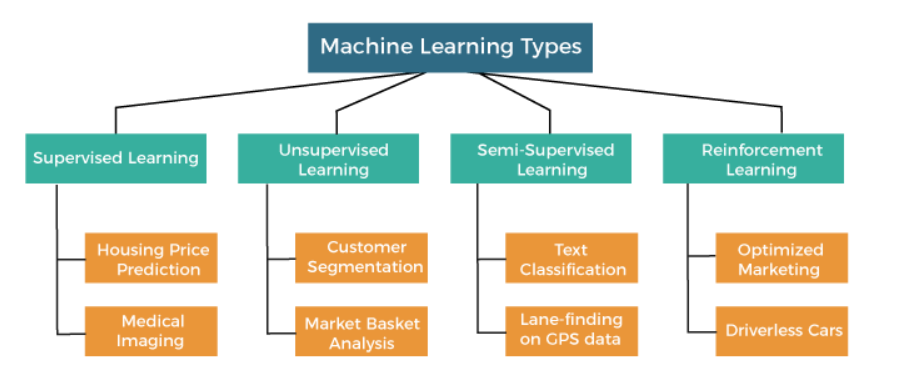
Ans: **Indirect learning** is a student-led learning process in which the lesson doesn't come directly from the teacher. Instead, it is student-centered. The key is to have students actively engaged in the learning process, making inquiries, using critical thinking skills to resolve problems, and testing hypotheses for validation. This high level of active student involvement strengthens the connections to the learned concepts.

iii. Self-education.

Ans: Self-learning AI is artificial intelligence that can train itself using unlabeled data. On a high level, it works by analyzing a dataset and looking for patterns that it can draw conclusions from. It essentially learns to “fill in the blanks.” Self-learning AI is especially useful when training a machine on a concept that does not have a lot of training data available. It can also come in handy for training computers on processes that researchers don’t know a lot about, making creating labeled training datasets difficult. Another benefit of self-learning AI is that once a new skill is learned, it can be more easily transferred to other similar skills. **Cybersecurity is one of the top areas where self-learning AI is currently being used, since it is better than most people at identifying changes and patterns indicating a breach.**

Q3: Provide a few examples of various types of machine learning.

Ans: The few examples of machine learning various types are:



Q4: Examine the various forms of machine learning.

Ans:The various forms of machine learning are:

1)**Supervised machine learning:**

Supervised learning is a type of machine learning that uses labeled data to train machine learning models. In labeled data, the output is already known. The model just needs to map the inputs to the respective outputs. They are basically useful for solving classification and regression problems such as weather forecasting, stock price analysis. Here some of the algorithms are listed as below:

* Linear Regression
* Logistics Regression
* K nearest neighbour
* Support vector machine
* Random forest
* Decision trees.
* Naïve Bayes.

2)**Unsupervised Machine learning:**

Unsupervised learning is a type of machine learning that uses unlabeled data to train machines. Unlabeled data doesn’t have a fixed output variable. The model learns from the data, discovers the patterns and features in the data, and returns the output. These are basically useful for solving clustering and association problems.Some of the Algorithms mentioned are:

* + K Means Clustering
  + Hierarchical Clustering
  + DBSCAN

3)**Reinforcement Machine Learning:**

Reinforcement Learning trains a machine to take suitable actions and maximize its rewards in a particular situation. It uses an agent and an environment to produce actions and rewards. In this learning technique, there is no predefined target variable. Reinforcement learning algorithms are widely used in the gaming industries to build games. It is also used to train robots to do human tasks.Some of the algorithms are :

* Q-Learning
* Sarsa
* Deep Q network

Q5: Can you explain what a well-posed learning problem is? Explain the main characteristics that must be present to identify a learning problem properly.

Ans: A well-posed problem can be defined as: “**A computer program is said to learn from experience E in context to some task T and some performance measure P, if its performance on T, as was measured by P, upgrades with experience E”.**

Any problem can be segregated as well-posed learning problem if it has three characteristics:

* Task:
* Performance Measure
* Experience

It can be more understood with the help of given example:

**A Robot Driving Problem**

* Task – driving on public four-lane highways using sight scanners
* Performance Measure – average distance progressed before a fallacy
* Experience – order of images and steering instructions noted down while observing a human drive

Q6: Is machine learning capable of solving all problems? Give a detailed explanation of your answer.

Ans: Most people think that they are familiar with machine learning and the relevant algorithms which used to classify or predict outcomes based on data. However, it is important to understand that machine learning is not the answer to all problems. Given the usefulness of machine learning, it can be hard to accept that sometimes it is not the best solution to a problem. There can be many limitations of machine learning such as:

1)**Ethics**: The idea of trusting data and algorithms more than our own judgment has its pros and cons. Obviously, we benefit from these algorithms, otherwise, we wouldn’t be using them in the first place. These algorithms allow us to automate processes by making informed judgments using available data. The most commonly discussed case currently is self-driving cars — how do we choose how the vehicle should react in the event of a fatal collision? In the future will we have to select which ethical framework we want our self-driving car to follow when we are purchasing the vehicle?

2)**Deterministic problem:** Running weather models is fine, but now that we have machine learning, can we just use this instead to obtain our weather forecasts? Can we leverage data from satellites, weather stations, and use an elementary predictive algorithm to discern whether it is going to rain tomorrow?

The answer is, surprisingly, yes. If we have knowledge of the air pressures around a certain region, the levels of moisture in the air, wind speeds, and information about neighboring points and their own variables, it becomes possible to train, for example, a neural network. But at what cost? **Machine learning is stochastic, not deterministic.** **A neural network does not understand Newton’s second law, or that density cannot be negative.**

3)**Lack of good data.**

Q7: What are the various methods and technologies for solving machine learning problems? Any two of them should be defined in detail.

Ans**:** For solving a machine learning problem there are various tools or technologies such as:

1) **Scikit Learn**.

Scikit-learn is for machine learning development in python. It provides a library for the Python programming language. It helps in data mining and data analysis.It provides models and algorithms for Classification, Regression, Clustering, Dimensional reduction, Model selection, and Pre-processing.

2)**Pytorch.**

Features:

* It helps in building neural networks through Autograd Module.
* It provides a variety of optimization algorithms for building neural networks.
* PyTorch can be used on cloud platforms.
* It provides distributed training, various tools, and libraries.

The methods which are opted for solving a machine learning model can be classified into two categories:

1)**Supervised machine learning**:  A supervised learning algorithm takes a known set of input data and known responses to the data (**output**) and trains a model to generate reasonable predictions for the response to the new data. Use supervised learning if we have known data for the output we are trying to estimate.It basically includes regression and classification problems.

2)**Unsupervised machine learning**:  It is used to eliminate datasets containing input data without labeled responses.

Clustering is a common unsupervised learning technique. It is used for exploratory data analysis to find hidden patterns and clusters in the data. Applications for cluster analysis include gene sequence analysis, market research, and commodity identification.

**For example**, if a cell phone company wants to optimize the locations where they build towers, they can use machine learning to predict how many people their towers are based on.

A phone can only talk to 1 tower at a time, so the team uses clustering algorithms to design the good placement of cell towers to optimize signal reception for their groups or groups of customers.

Common algorithms for performing clustering are **k-means** and **k-medoids, hierarchical clustering, Gaussian mixture models, hidden Markov models, self-organizing maps, fuzzy C-means clustering, and subtractive clustering.**

Q8: Can you explain the various forms of supervised learning? Explain each one with an example application.

Ans:The various forms forms of supervised learning are as follows:

**Supervised Learning**

**Regression**  **Classification**

**i)Linear Regression i)Logistic Regression**

**ii)SVR ii)SVM**

**iii)Decision Trees iii)Random Forest**

**iv)Naïve Bayes**

**1)Regression:** **In regression**, a single output value is produced using training data. This value is a probabilistic interpretation, which is ascertained after considering the strength of correlation among the input variables. For example, regression can help predict the price of a house based on its locality, size, etc.

**In logistic regression**, the output has discrete values based on a set of independent variables. This method can flounder when dealing with non-linear and multiple decision boundaries. Also, it is not flexible enough to capture complex relationships.

**2)Classification:** It involves grouping the data into classes.  If you are thinking of extending credit to a person, you can use classification to determine whether or not a person would be a loan defaulter. When the supervised learning algorithm labels input data into two distinct classes, it is called binary classification. Multiple classifications means categorizing data into more than two classes.

**3)Naïve bayesian model:** The Bayesian model of classification is used for large finite datasets. It is a method of assigning class labels using a direct acyclic graph. The graph comprises one parent node and multiple children nodes. And each child node is assumed to be independent and separate from the parent.

As the model for supervised learning in ML helps construct the classifiers in a simple and straightforward way, it works great with very small data sets. This model draws on common data assumptions, such as each attribute is independent. Yet having such simplification, this algorithm can easily be implemented on complex problems.

**4)Decision Trees:** A decision tree is a flowchart-like model that contains conditional control statements, comprising decisions and their probable consequences. The output relates to the labelling of unforeseen data.

In the tree representation, the leaf nodes correspond to class labels, and the internal nodes represent the attributes. A decision tree can be used to solve problems with discrete attributes as well as boolean functions. Some of the notable decision tree algorithms are ID3 and CART.

**5)Random Forest model**: The random forest model is an ensemble method. It operates by constructing a multitude of decision trees and outputs a classification of the individual trees.It basically gives the output as majority voting.

**6)Support vector machines(SVM):** SVM is highly popular amongst the supervised learning models as it can be used for classification or regression. Implementation of the model works well with high-dimensional spaces, but it can also be used effectively with small data sets. SVM can also classify new observations efficiently when the algorithm is trained on a data set. SVM performs this by creating singular or multiple hyperplanes to separate the data set between the two classes. The approach of segregation that SVM has makes it unique and more efficient among all the supervised learning models. However, analysing data with high dimensions can become problematic for this model. The simple reason is that SVM increases the dimensionality of the given data set to segregate it properly.

**7)Neural Networks:** This algorithm is designed to cluster raw input, recognize patterns, or interpret sensory data. Despite their multiple advantages, neural networks require significant computational resources. It can get complicated to fit a neural network when there are thousands of observations. It is also called the ‘black-box’ algorithm as interpreting the logic behind their predictions can be challenging.

Q9: What is the difference between supervised and unsupervised learning? With a sample application in each region, explain the differences.

Ans: The main distinction between the two approaches is the use of labeled datasets. To put it simply, supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not.The key differences can be as follows:

* **Goals:**In supervised learning, the goal is to predict outcomes for new data. You know up front the type of results to expect. With an unsupervised learning algorithm, the goal is to get insights from large volumes of new data. The machine learning itself determines what is different or interesting from the dataset.
* **Applications**: Supervised learning models are ideal for spam detection, weather forecasting and pricing predictions, among other things. In contrast, unsupervised learning is a great fit for anomaly detection, recommendation engines, customer personas and medical imaging.
* **Complexity:**Supervised learning is a simple method for machine learning, typically calculated through the use of programs like R or Python.In unsupervised learning, you need powerful tools for working with large amounts of unclassified data. Unsupervised learning models are computationally complex because they need a large training set to produce needed outcomes.

**Supervised Machine learning Algorithm (Example):**  Suppose we have an image of different types of fruits. The task of our supervised learning model is to identify the fruits and classify them accordingly. So to identify the image in supervised learning, we will give the input data as well as output for that, which means we will train the model by the shape, size, color, and taste of each fruit. Once the training is completed, we will test the model by giving the new set of fruit. The model will identify the fruit and predict the output using a suitable algorithm.

**Unsupervised Machine learning Algorithm (Example):**  To understand the unsupervised learning, we will use the example given above. So unlike supervised learning, here we will not provide any supervision to the model. We will just provide the input dataset to the model and allow the model to find the patterns from the data. With the help of a suitable algorithm, the model will train itself and divide the fruits into different groups according to the most similar features between them.

|  |  |
| --- | --- |
| **Supervised Learning** | **Unsupervised Learning** |
| Supervised learning algorithms are trained using labeled data. | Unsupervised learning algorithms are trained using unlabeled data. |
| Supervised learning model takes direct feedback to check if it is predicting correct output or not. | Unsupervised learning model does not take any feedback. |
| Supervised learning model predicts the output. | Unsupervised learning model finds the hidden patterns in data. |
| In supervised learning, input data is provided to the model along with the output. | In unsupervised learning, only input data is provided to the model. |
| The goal of supervised learning is to train the model so that it can predict the output when it is given new data. | The goal of unsupervised learning is to find the hidden patterns and useful insights from the unknown dataset. |
| Supervised learning needs supervision to train the model. | Unsupervised learning does not need any supervision to train the model. |
| Supervised learning can be categorized in **Classification** and **Regression** problems. | Unsupervised Learning can be classified in **Clustering** and **Associations** problems. |
| Supervised learning can be used for those cases where we know the input as well as corresponding outputs. | Unsupervised learning can be used for those cases where we have only input data and no corresponding output data. |
| Supervised learning model produces an accurate result. | Unsupervised learning model may give less accurate result as compared to supervised learning. |
| Supervised learning is not close to true Artificial intelligence as in this, we first train the model for each data, and then only it can predict the correct output. | Unsupervised learning is more close to the true Artificial Intelligence as it learns similarly as a child learns daily routine things by his experiences. |
| It includes various algorithms such as Linear Regression, Logistic Regression, Support Vector Machine, Multi-class Classification, Decision tree, Bayesian Logic, etc. | It includes various algorithms such as Clustering, KNN, and Apriori algorithm. |

Q10: Describe the machine learning process in depth.

Ans: Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://www.techtarget.com/whatis/definition/algorithm) use historical data as input to predict new output values. Some of the most common case include mail-spam detection, malware threat detection and predictive maintenance.Machine learning provides statistical tool to analyze,visualize, perform, prediction and other task with the help of data.

It basically works as we know that it basically uses two kind of techniques i.e **supervised learning**  which trains a model on known input and output data so that it can predict future outputs and **unsupervised learning** which finds hidden patterns or internal structures in input data. Basic steps include as :

1. Collecting Data.
2. Preparing and Visualizing the Data.
3. Choosing a Model.
4. Training the Model.
5. Evaluating the Model.
6. Hyperparameter Tuning.
7. Making Predictions.

Uses for machine learning include the following:

* **Customer relationship management.** [CRM software](https://www.techtarget.com/searchcustomerexperience/definition/CRM-customer-relationship-management) can use machine learning models to analyze email and prompt sales team members to respond to the most important messages first. More advanced systems can even recommend potentially effective responses.
* **Business intelligence.** [BI and analytics](https://www.techtarget.com/searchbusinessanalytics/definition/business-intelligence-BI) vendors use machine learning in their software to identify potentially important data points, patterns of data points and anomalies.
* **Human resource information systems.** [HRIS systems](https://www.techtarget.com/searchhrsoftware/definition/HRIS) can use machine learning models to filter through applications and identify the best candidates for an open position.
* **Self-driving cars.** Machine learning algorithms can even make it possible for a [semi-autonomous car](https://www.techtarget.com/searchenterpriseai/definition/driverless-car) to recognize a partially visible object and alert the driver.
* **Virtual assistants.** [Smart assistants](https://www.techtarget.com/searchcustomerexperience/definition/virtual-assistant-AI-assistant) typically combine supervised and unsupervised machine learning models to interpret natural speech and supply context.

a) Make brief notes on **any two** of the following:

i. MATLAB is one of the most widely used programming languages.

Ans: The full form of MATLAB is “Matrix Laboratory”   is a [proprietary](https://en.wikipedia.org/wiki/Proprietary_software) [multi-paradigm](https://en.wikipedia.org/wiki/Multi-paradigm_programming_language) [programming language](https://en.wikipedia.org/wiki/Programming_language) and [numeric computing](https://en.wikipedia.org/wiki/Numerical_analysis) environment developed by [MathWorks](https://en.wikipedia.org/wiki/MathWorks). MATLAB allows [matrix](https://en.wikipedia.org/wiki/Matrix_(mathematics)) manipulations, plotting of [functions](https://en.wikipedia.org/wiki/Function_(mathematics)) and data, implementation of [algorithms](https://en.wikipedia.org/wiki/Algorithm), creation of [user interfaces](https://en.wikipedia.org/wiki/User_interface), and interfacing with programs written in other languages. This programming language was designed for engineers and scientists to help them analyse and design products and systems that would help in the transformation of the world.

This programming language enables users to take ideas from research to production by deploying them in enterprise applications and embedded devices and integrating model-based design.

ii. Deep learning applications in healthcare.

Ans: Applications are as follows:

1. Drug discovery

Deep learning in healthcare helps in discovery of medicines and their development. The technology analyzes the patient's medical history and provides the best treatment for them. Moreover, this technology is gaining insights from patient symptoms and tests.

1. Medical imaging

Medical imaging techniques such as MRI scans, CT scans, ECG, are used to diagnose dreadful diseases such as heart disease, cancer, brain tumor. Hence, deep learning helps doctors to analyze the disease better and provide patients with the best treatment.

1. Insurance fraud

Deep learning is used to analyze the medical insurance fraud claims. With predictive analytics, it can predict fraud claims that are likely to happen in the future. Moreover, deep learning helps insurance industry to send out discounts and offers to their target patients.

1. Alzheimer's disease

Alzheimer is one of the significant challenges that medical industry faces. Deep learning technique is used to detect Alzheimer’s disease at an early stage.

1. Genome

Deep learning technique is used to understand a genome and help patients get an idea about disease that might affect them.

iii. Study of the market basket.

Ans:xxxxxxxxxxxx

iv. Linear regression (simple).

Ans: Simple linear regression is a statistical method for establishing the relationship between two variables using a straight line. The line is drawn by finding the slope and intercept, which define the line and minimize regression errors.

The simplest form of simple linear regression has only one x variable and one y variable. The x variable is the independent variable because it is independent of what you try to predict the dependent variable. The y variable is the dependent variable because it depends on what you try to predict.

**y = β0 +β1x+ε** is the formula used for simple linear regression.

* y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).
* B0 is the intercept, the predicted value of y when the x is 0.
* B1 is the regression coefficient how much we expect y to change as x increases.
* x is the independent variable (the variable we expect is influencing y).
* e is the error of the estimate, or how much variation there is in our regression coefficient estimate.

Simple linear regression establishes a line that fits your data, but it does not guarantee that the line is good enough. For example, if our data points have an upward trend and are very far apart, then simple linear regression will give us a downward-sloping line, which will not match our data.

Q11:Make a comparison between:-

1. Generalization and abstraction.

Ans: **Abstraction:** It involves the translation of data into broader representations. Representing raw input data in a structured format is the very essential task for a learning algorithm. Prior to this point, the data is merely ones and zeros on a disk or in memory, they have no meaning. The work of assigning a meaning to data occurs during the **abstraction** process.

**Generalization:** The term generalization describes the process of turning abstracted knowledge into a form that can be utilized for action. Generalization is a somewhat vague process that is a bit difficult to describe. Traditionally, it has been imagined as a search through the entire set of models (that is, theories) that could have been abstracted during training. Specifically, if we imagine a hypothetical set containing every possible theory that could be established from the data, generalization involves the reduction of this set into a manageable number of important findings.

2. Learning that is guided and unsupervised.

Ans: **Unsupervised machine learning:** Unsupervised learning is a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision. Unsupervised learning cannot be directly applied to a regression or classification problem because unlike supervised learning, we have the input data but no corresponding output data. The goal of unsupervised learning is to find the underlying structure of dataset, group that data according to similarities, and represent that dataset in a compressed format.

**Guided learning:** Human-guided machine learning is a process whereby subject matter experts accelerate the learning process by teaching the technology in real-time. For example, if the machine learning model comes across a piece of data it is uncertain about, a human can be asked to weigh in and give feedback. The model then learns from this input, and uses it to make a more accurate prediction the next time. Human-guided machine learning works from the bottom up by first using algorithms to conduct the heavy lifting of identifying relationships within the data, and engaging humans when necessary for training or validation. This means that, inevitably, the amount of time a human needs to spend performing a specific task will decrease as the machine learning accuracy increases.

3. Regression and classification.

Ans: Regression and Classification algorithms are Supervised Learning algorithms. Both the algorithms are used for prediction in Machine learning and work with the labeled datasets. But the difference between both is how they are used for different machine learning problems.

The main difference between Regression and Classification algorithms that Regression algorithms are used to predict the continuous values such as price, salary, age, etc. and Classification algorithms are used to predict/Classify the discrete values such as Male or Female, True or False, Spam or Not Spam, etc.

**Types of ML Classification Algorithms:**

Classification Algorithms can be further divided into the following types:

* Logistic Regression
* K-Nearest Neighbours
* Support Vector Machines
* Kernel SVM
* Naïve Bayes
* Decision Tree Classification
* Random Forest Classification

**Types of Regression Algorithm:**

* Simple Linear Regression
* Multiple Linear Regression
* Polynomial Regression
* Support Vector Regression
* Decision Tree Regression
* Random Forest Regression