**UNIT-1(part-1)**

**1a. Define Machine Learning. Why Machine Learning is important?**

Ans **: What Is Machine Learning:**

Machine Learning is the science (and art) of programming computers so they can learn from data.

Here is a slightly more general definition:

**Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed**.—Arthur Samuel, 1959

And a more engineering-oriented one: **A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.**—Tom Mitchell, 1997

Your spam filter is a Machine Learning program that, given examples of spam emails (e.g., flagged by users) and examples of regular (non-spam, also called “ham”) emails, can learn to flag spam. The examples that the system uses to learn are called the training set. Each training example is called a training instance (or sample). In this case, the task T is to flag spam for new emails, the experience E is the training data, and the performance measure P needs to be defined; for example, you can use the ratio of correctly classified emails. This particular performance measure is called accuracy,and it is often used in classification tasks.

**Why Machine Learning is important:**

Machine Learning can help humans learn . ML algorithms can be inspected to see what they have learned (although for some algorithms this can be tricky). For instance, once a spam filter has been trained on enough spam, it can easily be inspected to reveal the list of words and combinations of words that it believes are the best predictors of spam. Sometimes this will reveal unsuspected correlations or new trends, and thereby lead to a better understanding of the problem. Applying ML techniques to dig into large amounts of data can help discover patterns that were not immediately apparent. This is called data mining.

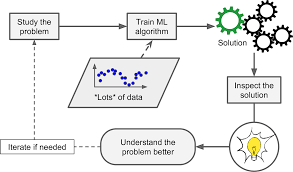


Figure : Machine Learning can help humans learn

To summarize, Machine Learning is important for:

• Problems for which existing solutions require a lot of fine-tuning or long lists ofrules: one Machine Learning algorithm can often simplify code and perform bet‐ter than the traditional approach.

• Complex problems for which using a traditional approach yields no good solu‐tion: the best Machine Learning techniques can perhaps find a solution.

• Fluctuating environments: a Machine Learning system can adapt to new data.

• Getting insights about complex problems and large amounts of data.

**1b. Differentiate traditional programming Vs machine learning?**

|  |  |
| --- | --- |
| **Traditional Programming** | **Machine Learning** |
| In traditional programming, rule-based code is written by the developers depending on the problem statements. | Machine Learning is a subset of artificial intelligence(AI) that focus on learning from data to develop an algorithm that can be used to make a prediction. |
| Traditional programming is typically rule-based and deterministic. It hasn’t self-learning features like Machine Learning and AI. | Machine Learning uses a data-driven approach, It is typically trained on historical data and then used to make predictions on new data. |
| Traditional programming is totally dependent on the intelligence of developers. So, it has very limited capability. | ML can find patterns and insights in large datasets that might be difficult for humans to discover. |
| The input data, and someone (programmer) coded a program that uses that data and runs on a computer to produce the desired output.  https://miro.medium.com/v2/resize:fit:1050/1*TPT1fHu5KpE3vkl_EGYyNw.png | The input data and output are fed to an algorithm to create a program.  https://miro.medium.com/v2/resize:fit:1050/1*_xv__NEFZwjP1KZz-nisrw.png |
| In Traditional programming one has to manually formulate/code rules | In Machine Learning the algorithms automatically formulate the rules from the data, which is very powerful. |
| Traditional programming is a more fixed approach where the programmer designs the solution explicitly | ML is a more flexible and adaptive approach where the ML model learns from data to generate a solution. |
| Traditional programming is often used to build applications and software systems that have specific functionality. | Machine Learning is the subset of AI. And Now it is used in various AI-based tasks like Chatbot Question answering, self-driven car., etc |

**2. Discuss some applications of machine learning with examples?**

Machine learning is a buzzword for today's technology, and it is growing very rapidly day by day. We are using machine learning in our daily life even without knowing it such as Google Maps, Google assistant, Alexa, etc. Below are some most trending real-world applications of Machine Learning:



1. Image Recognition:

Image recognition is one of the most common applications of machine learning. It is used to identify objects, persons, places, digital images, etc. The popular use case of image recognition and face detection is, **Automatic friend tagging suggestion**:

Facebook provides us a feature of auto friend tagging suggestion. Whenever we upload a photo with our Facebook friends, then we automatically get a tagging suggestion with name, and the technology behind this is machine learning's **face detection** and **recognition algorithm**.

It is based on the Facebook project named "**Deep Face**," which is responsible for face recognition and person identification in the picture.

2. Speech Recognition

While using Google, we get an option of "**Search by voice**," it comes under speech recognition, and it's a popular application of machine learning.

Speech recognition is a process of converting voice instructions into text, and it is also known as "**Speech to text**", or "**Computer speech recognition**." At present, machine learning algorithms are widely used by various applications of speech recognition. **Google assistant**, **Siri**, **Cortana**, and **Alexa** are using speech recognition technology to follow the voice instructions.

3. Traffic prediction:

If we want to visit a new place, we take help of Google Maps, which shows us the correct path with the shortest route and predicts the traffic conditions.

It predicts the traffic conditions such as whether traffic is cleared, slow-moving, or heavily congested with the help of two ways:

* **Real Time location** of the vehicle form Google Map app and sensors
* **Average time has taken** on past days at the same time.

Everyone who is using Google Map is helping this app to make it better. It takes information from the user and sends back to its database to improve the performance.

4. Product recommendations:

Machine learning is widely used by various e-commerce and entertainment companies such as **Amazon**, **Netflix**, etc., for product recommendation to the user. Whenever we search for some product on Amazon, then we started getting an advertisement for the same product while internet surfing on the same browser and this is because of machine learning.

Google understands the user interest using various machine learning algorithms and suggests the product as per customer interest.

As similar, when we use Netflix, we find some recommendations for entertainment series, movies, etc., and this is also done with the help of machine learning.

5. Self-driving cars:

One of the most exciting applications of machine learning is self-driving cars. Machine learning plays a significant role in self-driving cars. Tesla, the most popular car manufacturing company is working on self-driving car. It is using unsupervised learning method to train the car models to detect people and objects while driving.

6. Email Spam and Malware Filtering:

Whenever we receive a new email, it is filtered automatically as important, normal, and spam. We always receive an important mail in our inbox with the important symbol and spam emails in our spam box, and the technology behind this is Machine learning. Below are some spam filters used by Gmail:

* Content Filter
* Header filter
* General blacklists filter
* Rules-based filters
* Permission filters

Some machine learning algorithms such as **Multi-Layer Perceptron**, **Decision tree**, and **Naïve Bayes classifier** are used for email spam filtering and malware detection.

7. Virtual Personal Assistant:

We have various virtual personal assistants such as **Google assistant**, **Alexa**, **Cortana**, **Siri**. As the name suggests, they help us in finding the information using our voice instruction. These assistants can help us in various ways just by our voice instructions such as Play music, call someone, Open an email, Scheduling an appointment, etc.

These virtual assistants use machine learning algorithms as an important part.

These assistant record our voice instructions, send it over the server on a cloud, and decode it using ML algorithms and act accordingly.

8. Online Fraud Detection:

Machine learning is making our online transaction safe and secure by detecting fraud transaction. Whenever we perform some online transaction, there may be various ways that a fraudulent transaction can take place such as **fake accounts**, **fake ids**, and **steal money** in the middle of a transaction. So to detect this, **Feed Forward Neural network** helps us by checking whether it is a genuine transaction or a fraud transaction.

For each genuine transaction, the output is converted into some hash values, and these values become the input for the next round. For each genuine transaction, there is a specific pattern which gets change for the fraud transaction hence, it detects it and makes our online transactions more secure.

9. Stock Market trading:

Machine learning is widely used in stock market trading. In the stock market, there is always a risk of up and downs in shares, so for this machine learning's **long short term memory neural network** is used for the prediction of stock market trends.

10. Medical Diagnosis:

In medical science, machine learning is used for diseases diagnoses. With this, medical technology is growing very fast and able to build 3D models that can predict the exact position of lesions in the brain.

It helps in finding brain tumors and other brain-related diseases easily.

11. Automatic Language Translation:

Nowadays, if we visit a new place and we are not aware of the language then it is not a problem at all, as for this also machine learning helps us by converting the text into our known languages. Google's GNMT (Google Neural Machine Translation) provide this feature, which is a Neural Machine Learning that translates the text into our familiar language, and it called as automatic translation.

The technology behind the automatic translation is a sequence to sequence learning algorithm, which is used with image recognition and translates the text from one language to another language.

**3. Explain different perspectives and Issues in machine learning?**

**Perspectives in ML**:

Machine Learning involves searching strategy in a very large hypothesis space to determine single solution that best fits the observed data with any prior knowledge of the learner.

Checkers learner example:

Task T: playing checkers

Performance measure P: Percent of games won in the world tournament

Training experience E: Games played against itself

The hypothesis space consists of all evaluation functions that can be represented by some choice of values for the weights Wo through W

Task: search through this vast space to locate the hypothesis that is most consistent with the available

training examples.

• The Least Mean Squares (LMS) algorithm for fitting weights to achieve the goal by iteratively tuning the weights, adding a correction to each weight each time the hypothesized evaluation function predicts a value that differs from the training value.

• Hypothesis space represented by linear functions, logical descriptions, decision trees, artificial neural networks.

**Issues in machine learning**:

In short, since your main task is to select a learning algorithm and train it on some

data, the two things that can go wrong are “bad algorithm” and “bad data.” Let’s start

with examples of bad data.

**Insufficient Quantity of Training Data:**

For a toddler to learn what an apple is, all it takes is for you to point to an apple and say “apple” (possibly repeating this procedure a few times). Now the child is able to recognize apples in all sorts of colors and shapes. Genius

Machine Learning is not quite there yet; it takes a lot of data for most Machine Learning algorithms to work properly. Even for very simple problems you typically need thousands of examples, and for complex problems such as image or speech recognition you may need millions of examples (unless you can reuse parts of an existing model).

**Non-representative Training Data:**

In order to generalize well, it is crucial that your training data be representative of the new cases you want to generalize to. This is true whether you use instance-based learning or model-based learning.

For example, the set of countries we used earlier for training the linear model was not perfectly representative; a few countries were missing. Figure 1-21 shows what the data looks like when you add the missing countries.

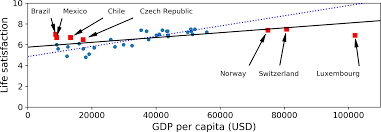


Figure : A more representative training sample

**Poor-Quality Data:**

Obviously, if your training data is full of errors, outliers, and noise (e.g., due to poorquality measurements), it will make it harder for the system to detect the underlying patterns, so your system is less likely to perform well. It is often well worth the effort to spend time cleaning up your training data. The truth is, most data scientists spend a significant part of their time doing just that. The following are a couple of examples of when you’d want to clean up training data:

• If some instances are clearly outliers, it may help to simply discard them or try to fix the errors manually.

• If some instances are missing a few features (e.g., 5% of your customers did not specify their age), you must decide whether you want to ignore this attribute altogether, ignore these instances, fill in the missing values (e.g., with the median age), or train one model with the feature and one model without it.

**Irrelevant Features:**

As the saying goes: garbage in, garbage out. Your system will only be capable of learning if the training data contains enough relevant features and not too many irrelevant ones. A critical part of the success of a Machine Learning project is coming up with a good set of features to train on. This process, called feature engineering, involves the following steps:

• Feature selection (selecting the most useful features to train on among existing features)

• Feature extraction (combining existing features to produce a more useful one—as we saw earlier, dimensionality reduction algorithms can help)

• Creating new features by gathering new data Now that we have looked at many examples of bad data,

let’s look at a couple of examples of bad algorithms:

**Overfitting the Training Data :**

Say you are visiting a foreign country and the taxi driver rips you off. You might be tempted to say that all taxi drivers in that country are thieves. Overgeneralizing is something that we humans do all too often, and unfortunately machines can fall into the same trap if we are not careful. In Machine Learning this is called overfitting: it means that the model performs well on the training data, but it does not generalize well.

Figure shows an example of a high-degree polynomial life satisfaction model that strongly overfits the training data. Even though it performs much better on the training data than the simple linear model, would you really trust its predictions?

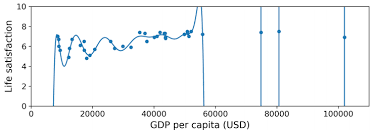


Figure. Overfitting the training data

**Underfitting the Training Data:**

As you might guess, underfitting is the opposite of overfitting: it occurs when your model is too simple to learn the underlying structure of the data. For example, a linear model of life satisfaction is prone to underfit; reality is just more complex than the model, so its predictions are bound to be inaccurate, even on the training examples.

Here are the main options for fixing this problem:

• Select a more powerful model, with more parameters.

• Feed better features to the learning algorithm (feature engineering).

• Reduce the constraints on the model (e.g., reduce the regularization hyperparameter).

**4. what do you mean by well posed learning problem? Differentiate supervised, unsupervised, and reinforcement learning?**

**Well Posed Learning Problem –** 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 traits –

* Task
* Performance Measure
* Experience

**Certain examples that efficiently defines the**well-posed**learning problem are –**

**1. To better filter emails as spam or not**

* Task – Classifying emails as spam or not
* Performance Measure – The fraction of emails accurately classified as spam or not spam
* Experience – Observing you label emails as spam or not spam

**2. A checkers learning problem**

* Task – Playing checkers game
* Performance Measure – percent of games won against opposer
* Experience**–** playing implementation games against itself

**3. Handwriting Recognition Problem**

* Task – Acknowledging handwritten words within portrayal
* Performance Measure – percent of words accurately classified
* Experience – a directory of handwritten words with given classifications

**4. 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 driver

**5. Fruit Prediction Problem**

* Task – forecasting different fruits for recognition
* Performance Measure – able to predict maximum variety of fruits
* Experience – training machine with the largest datasets of fruits images

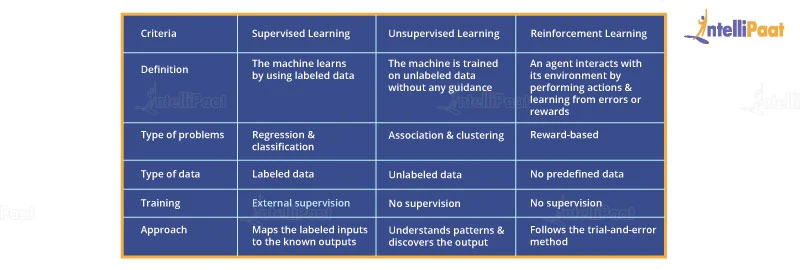
**6. Face Recognition Problem**

* Task – predicting different types of faces
* Performance Measure – able to predict maximum types of faces
* Experience – training machine with maximum amount of datasets of different face images

**7. Automatic Translation of documents**

* Task – translating one type of language used in a document to other language
* Performance Measure – able to convert one language to other efficiently
* Experience – training machine with a large dataset of different types of languages

Differentiate supervised, unsupervised, and reinforcement learning?



**5. Differentiate Artificial Intelligence vs Deep Learning vs machine learning with suitable example?**

| **Artificial Intelligence (AI)** | **Machine Learning (ML)** | **Deep Learning (DL)** |
| --- | --- | --- |
| AI simulates human intelligence to perform tasks and make decisions. | ML is a subset of AI that uses algorithms to learn patterns from data. | DL is a subset of ML that employs artificial neural networks for complex tasks. |
| AI may or may not require large datasets; it can use predefined rules. | ML heavily relies on labeled data for training and making predictions. | DL requires extensive labeled data and performs exceptionally with big datasets. |
| AI can be rule-based, requiring human programming and intervention. | ML automates learning from data and requires less manual intervention. | DL automates feature extraction, reducing the need for manual engineering. |
| AI can handle various tasks, from simple to complex, across domains. | ML specializes in data-driven tasks like classification, regression, etc. | DL excels at complex tasks like image recognition, natural language processing, and more. |
| AI algorithms can be simple or complex, depending on the application. | ML employs various algorithms like decision trees, SVM, and random forests. | DL relies on deep neural networks, which can have numerous hidden layers for complex learning. |
| AI may require less training time and resources for rule-based systems. | ML training time varies with the algorithm complexity and dataset size. | DL training demands substantial computational resources and time for deep networks. |
| AI systems may offer interpretable results based on human rules. | ML models can be interpretable or less interpretable based on the algorithm. | DL models are often considered less interpretable due to complex network architectures. |
| AI is used in virtual assistants, recommendation systems, and more. | ML is applied in image recognition, spam filtering, and other data tasks. | DL is utilized in autonomous vehicles, speech recognition, and advanced AI applications. |

