

**BOWEN UNIVERSITY, IWO**

**COLLEGE OF COMPUTING AND COMMUNICATION STUDIES**

**DEPARTMENT OF COMPUTER SCIENCE**

**CSC 404- MACHINE LEARNING WITH PYTHON**

**LECTURE NOTE 4**

Methods for Machine Learning

There are various ML algorithms, techniques and methods that can be used to build models for solving real-life problems by using data. In this chapter, we are going to discuss such different kinds of methods.

**Different Types of Methods**

****

The following are various ML methods based on some broad categories:

**Based on human supervision**

In the learning process, some of the methods that are based on human supervision are as follows:

**Supervised Learning**

Supervised learning algorithms or methods are the most commonly used ML algorithms. This method or learning algorithm take the data sample i.e. the training data and its associated output i.e. labels or responses with each data samples during the training process.

The main objective of supervised learning algorithms is to learn an association between input data samples and corresponding outputs after performing multiple training data instances.

For example, we have

1. Input variables and
2. Output variable

Now, apply an algorithm to learn the mapping function from the input to output as follows:

Y=f(x)

Now, the main objective would be to approximate the mapping function so well that even when we have new input data (x), we can easily predict the output variable (Y) for that new input data.

It is called supervised because the whole process of learning can be thought as it is being supervised by a teacher or supervisor. Examples of supervised machine learning algorithms includes **Decision tree, Random Forest, KNN, Logistic Regression** etc.

Based on the ML tasks, supervised learning algorithms can be divided into following two broad classes:

* Classification
* Regression

**Classification**

The key objective of classification-based tasks is to predict categorial output labels or responses for the given input data. The output will be based on what the model has learned in training phase. As we know that the categorial output responses means unordered and discrete values, hence each output response will belong to a specific class or category. We will discuss Classification and associated algorithms in detail in the upcoming chapters also.

**Regression**

The key objective of regression-based tasks is to predict output labels or responses which are continues numeric values, for the given input data. The output will be based on what the model has learned in its training phase. Basically, regression models use the input data features (independent variables) and their corresponding continuous numeric output values (dependent or outcome variables) to learn specific association between inputs and corresponding outputs. We will discuss regression and associated algorithms in detail in further chapters also.

**Unsupervised Learning**

As the name suggests, it is opposite to supervised ML methods or algorithms which means in unsupervised machine learning algorithms we do not have any supervisor to provide any sort of guidance. Unsupervised learning algorithms are handy in the scenario in which we do not have the liberty, like in supervised learning algorithms, of having pre-labeled training data and we want to extract useful pattern from input data.

For example, it can be understood as follows:

Suppose we have:

1. **Input variables**, then there would be no corresponding output variable and the algorithms need to discover the interesting pattern in data for learning.

Examples of unsupervised machine learning algorithms includes **K-means clustering, K-nearest neighbors** etc.

Based on the ML tasks, unsupervised learning algorithms can be divided into following broad classes:

* Clustering
* Association
* Dimensionality Reduction

**Clustering**

Clustering methods are one of the most useful unsupervised ML methods. These algorithms used to find similarity as well as relationship patterns among data samples and then cluster those samples into groups having similarity based on features. The real-world example of clustering is to group the customers by their purchasing behavior.

**Association**

Another useful unsupervised ML method is **Association** which is used to analyze large dataset to find patterns which further represents the interesting relationships between various items. It is also termed as **Association Rule Mining** or **Market basket analysis** which is mainly used to analyze customer shopping patterns.

**Dimensionality Reduction**

This unsupervised ML method is used to reduce the number of feature variables for each data sample by selecting set of principal or representative features. A question arises here is that why we need to reduce the dimensionality? The reason behind is the problem of feature space complexity which arises when we start analyzing and extracting millions of features from data samples. This problem generally refers to “curse of dimensionality”. PCA (Principal Component Analysis), K-nearest neighbors and discriminant analysis are some of the popular algorithms for this purpose.

**Anomaly Detection**

This unsupervised ML method is used to find out the occurrences of rare events or observations that generally do not occur. By using the learned knowledge, anomaly detection methods would be able to differentiate between anomalous or a normal data point. Some of the unsupervised algorithms like clustering, KNN can detect anomalies based on the data and its features.

**Semi-supervised Learning**

Such kind of algorithms or methods are neither fully supervised nor fully unsupervised. They basically fall between the two i.e. supervised and unsupervised learning methods. These kinds of algorithms generally use small supervised learning component i.e. small amount of pre-labeled annotated data and large unsupervised learning component i.e. lots of unlabeled data for training. We can follow any of the following approaches for implementing semi-supervised learning methods:

* The first and simple approach is to build the supervised model based on small amount of labeled and annotated data and then build the unsupervised model by applying the same to the large amounts of unlabeled data to get more labeled samples. Now, train the model on them and repeat the process.
* The second approach needs some extra efforts. In this approach, we can first use the unsupervised methods to cluster similar data samples, annotate these groups and then use a combination of this information to train the model.

**Reinforcement Learning**

These methods are different from previously studied methods and very rarely used also. In this kind of learning algorithms, there would be an agent that we want to train over a period of time so that it can interact with a specific environment. The agent will follow a set of strategies for interacting with the environment and then after observing the environment it will take actions regards the current state of the environment. The following are the main steps of reinforcement learning methods:

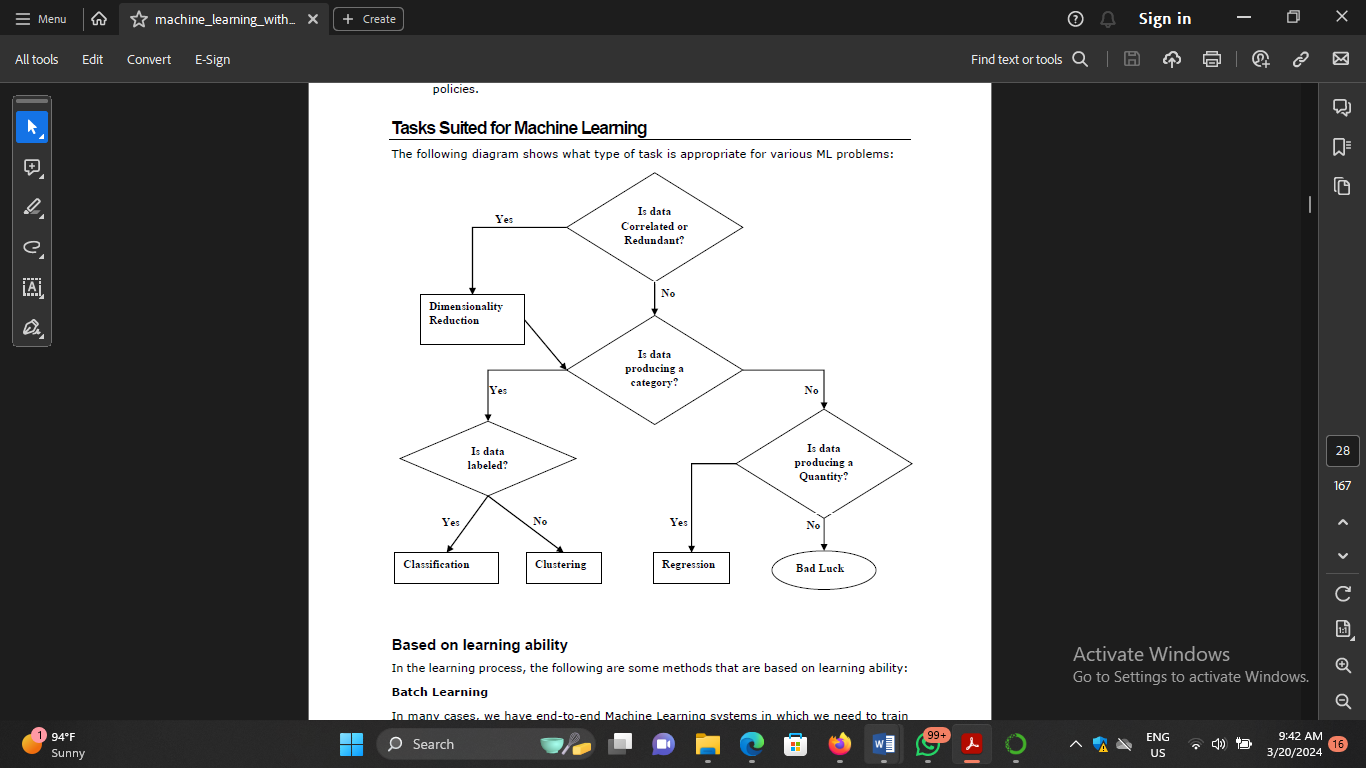
* **Step1:** First, we need to prepare an agent with some initial set of strategies.
* **Step2:** Then observe the environment and its current state.
* **Step3:** Next, select the optimal policy regards the current state of the environment and perform important action.
* **Step4:** Now, the agent can get corresponding reward or penalty as per accordance with the action taken by it in previous step.

* **Step5:** Now, we can update the strategies if it is required so.
* **Step6:** At last, repeat steps 2-5 until the agent got to learn and adopt the optimal policies.

**Tasks Suited for Machine Learning**

****

The following diagram shows what type of task is appropriate for various ML problems:



**Based on learning ability**

In the learning process, the following are some methods that are based on learning ability:

**Batch Learning**

In many cases, we have end-to-end Machine Learning systems in which we need to train the model in one go by using whole available training data. Such kind of learning method or algorithm is called **Batch or Offline learning**. It is called Batch or Offline learning because it is a one-time procedure and the model will be trained with data in one single batch. The following are the main steps of Batch learning methods:

**Step1:** First, we need to collect all the training data for start training the model.

**Step2:** Now, start the training of model by providing whole training data in one go.

**Step3:** Next, stop learning/training process once you got satisfactory results/performance.

**Step4:** Finally, deploy this trained model into production. Here, it will predict the output for new data sample.

**Online Learning**

It is completely opposite to the batch or offline learning methods. In these learning methods, the training data is supplied in multiple incremental batches, called mini-batches, to the algorithm. Followings are the main steps of Online learning methods:

**Step1:** First, we need to collect all the training data for starting training of the model.

**Step2:** Now, start the training of model by providing a mini-batch of training data to the algorithm.

**Step3:** Next, we need to provide the mini-batches of training data in multiple increments to the algorithm.

**Step4:** As it will not stop like batch learning hence after providing whole training data in mini-batches, provide new data samples also to it.

**Step5:** Finally, it will keep learning over a period of time based on the new data samples.

**Based on Generalization Approach**

In the learning process, followings are some methods that are based on generalization approaches:

**Instance based Learning**

Instance based learning method is one of the useful methods that build the ML models by doing generalization based on the input data. It is opposite to the previously studied learning methods in the way that this kind of learning involves ML systems as well as methods that uses the raw data points themselves to draw the outcomes for newer data samples without building an explicit model on training data.

In simple words, instance-based learning basically starts working by looking at the input data points and then using a similarity metric, it will generalize and predict the new data points.

**Model based Learning**

In Model based learning methods, an iterative process takes place on the ML models that are built based on various model parameters, called hyperparameters and in which input data is used to extract the features. In this learning, hyperparameters are optimized based on various model validation techniques. That is why we can say that Model based learning methods uses more traditional ML approach towards generalization.