Problem Set 1

Name: Rupali Roy (EID: rur73)

1. Artificial Intelligence versus Machine Learning [4 points]: In your own words, define artificial intelligence and define machine learning. Then describe the relationship between the two disciplines.

Ans. **Artificial Intelligence** is a concept in which allows us to create intelligent machines that can simulate human thinking capability and behavior.

Machine learning on other hand uses statistical learning algorithms to build systems that have the ability to automatically learn and improve from experiences without being explicitly programmed. Here, we train the algorithm by providing it with a lot of data so that they can give accurate output.

The main difference is that Machine Learning is a subset of AI. In AI we make intelligent systems to perform any task like a human whereas in ML we teach machines with data to perform a particular task and give an accurate result.

2. Supervised versus Unsupervised Learning [4 points]: In your own words, define each learning approach and describe the difference between these approaches.

Ans. **Supervised Learning:**

As the name suggests supervised learning can be defined as a machine learning technique in which the model is trained in the presence of a supervisor(labels). In this technique we use a labelled dataset, we have input variables (x) and an output variable (Y) and we use different algorithms to learn the data mapping from input to output. After training the model with labelled dataset, we test the model's accuracy by generating prediction for the test dataset. The test dataset is a new dataset which was not used to train the model.

Unsupervised Learning is a machine learning technique which is used in scenarios where we do not have labelled dataset. In this the machine learns more about the data by identifying and inferring different patterns in the dataset. It is called unsupervised since the algorithms are independent and are left on their own to group the data based on similarities, differences, and patterns. It is mostly used for clustering and dimensionality reduction.

The main difference between both types of learning is the type of data that is used. For supervised as described above we use labelled dataset and for unsupervised we use unlabeled dataset. Unsupervised algorithms try to identify patterns in the data and then groups similar data.

3. Supervised Learning: Regression versus Classification [4 points]: In your own words, define each learning problem and describe the difference between them.

In both Regression and Classification problem the machine learning algorithm is trained using labelled dataset.

Regression: In case of regression, the main aim of the machine learning algorithm is to predict the continuous target value(numeric). For e.g., Predicting the salary of a person, given the features like past work experience, type of degree, years of relevant work experience etc. In this we mathematically

estimate a mapping function (f) from the input variables (x-features) to output variables (y). The regression algorithm is evaluated by using the root mean squared error of its ouput.

Classification: A classification algorithm as the name suggests is used to create a model that classifies the input data into discrete classes or labels. A classification problem can have target variable which is ordinal; the data is classified into orders for e.g. discrete university ranking from 1 to 5 or the target variable can be categorical e.g. the data is classified into different categories (A, B or C).

The main difference between regression and classification is the type of predictor variable in both the algorithms. The regression algorithm is used to predict a target numeric/continuous variable whereas classification algorithm is used to predict discrete value. Whenever we want to classify a data point into different classes or levels we use classification algorithm and whenever we need to predict a continuous numerical value we use regression algorithm.

4. Supervised Learning: Offline versus Online Learning [4 points]: In your own words, define each learning approach and describe the difference between them.

Ans.

Offline learning is used in scenarios where a system is not able to learn incrementally or on the fly. In such cases we first train the model using all the available data. Since the data quantity is huge it usually takes hours for model training and hence we do it offline. After the model is trained on the dataset, we launch it in the production environment and where it is used to make predictions and it runs without any further training.

Online Training: In online learning the data is coming to the model sequentially (individually or in batches) and the model needs to learn from it in a sequential manner. The system learns in an incremental way by using the input data instances. It is mostly used in scenarios where a lot of real time continuous flow of data happens for e.g. stock price prediction.

The main difference would be the way the model learns from the data, in offline learning all the training happens before the model is launched in production. The model is trained on the entire available dataset. Whereas in online training the model is initially trained on a dataset and after launching it continues to train itself on the fly, by using the input data streams and continuously uses the data instances to train the model.

5. Instance-Based versus Model-Based Learning [4 points]: In your own words, define each learning approach and describe the difference between them.

Ans. In machine learning in order to check the model performance it is really important to consider how the model generalizes. By generalization we mean how well a ML model performs on a new set of data rather than just the data it was trained on.

Instance Based Learning: This type of learning can also be referred as memory based learning, here the system just learns the examples that were given during the training, it does not perform explicit generalization, whenever a new data instance is encountered it tries to compare it with the instances that were present during the training and generalizes it using a similarity measure.

Model Based Learning algorithms use the training data to create a machine learning model that has a set of parameters learned from the training data. We try to tune these parameters in order to get accurate results of the prediction.

The difference between model based and instance based is the absence of parameters. In instance based the system just by hearts the different training examples. In model based, we try to build a model and train the model using a dataset, we tune the parameters of the model to get best accuracy.