**DAY 4 [19/01/2022]: Online Machine Learning**

**2.Online Learning**

**Have you ever heard of the companies who says, more the usage of our products, the performance improves as such. Based on the usability, the performance improves. It is nothing but online machine learning. Dynamically on the server, ml model will train with the new data and improves.**

**Online Learning unlike batch learning, training is done incrementally. Sequentially (small batches) the model is fed with the data. After each training, the model improves. Since these batches are small chunk of data, training would be done on server as in with the production. This is why it is called as Online Learning.**

**Initially we start with the small data and feed it to the ml model. Once it is trained, testing will be done to check is it working or not. Then it will be passed on to the server. On the Server, there is a continuous inflow of new data.**

**But the model performs two functions,**

**1. Prediction**

**2. Learning.**

**That’s the main difference between online and batch and learning. In batch learning, once the training is done, it will be deployed on to the server. On server it does only prediction. If we need to train it again, make it offline and train. Whereas in online learning, model can be trained dynamically on the server. As in new data comes in, model will keep on improving.**

**Example: Chatbots like google, alexa**

**Youtube – by clicking a video, new data would be generated and retrained with the data which results in recommendation of other similar videos.**

* **When to use?**

1. **Where there is a concept drift**

**Sometimes we make a ml model to solve a problem where its nature is volatile in nature. Example: stocks, ecommerce platforms**

1. **Cost effective**

**Since we work on small batches of data.**

1. **Faster Solution**

* **How to Implement?**

**We can use scikit learn library**

**Or can also use River Library** [**https://github.com/online-ml/river**](https://github.com/online-ml/river)

[**https://riverml.xyz/dev/**](https://riverml.xyz/dev/)

**Or can also use Vompal Wabbit Library**

[**https://vowpalwabbit.org/**](https://vowpalwabbit.org/)

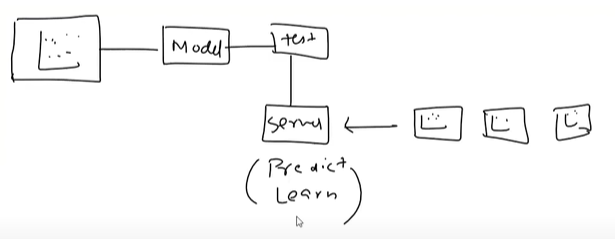
* **Learning rate**

**How frequently we train the data.**

**We cannot train as in when comes new data every time. This results in lose of old data. But slowly taking in is also not efficient. Hence there must be correct running rate which remembers old data and also take in new data as well. Or else the model would misbehave and won’t work well.**

* **Out of core running**

**The dataset will be so large such that we cannot train it at a time using batch learning. The data is so huge that we cannot load it at once in the memory. If isn’t possible to load then how to train? In such cases we use online learning.**



**Hence what we do is covert large dataset into small datasets and then load and train it one by one. We d this offline but the technique is Online.**

* **Disadvantage**

1. **Tricky to use**

**Training the model isn’t a problem. But expecting results from the model as we need is a difficult one due to learning rate, processing data on server, handling with data speed.**

1. **Risky**

**If some wrong data go on to server by a spammer or someone else, server hack then the model becomes biased. The solution is having an active monitoring system or else off the server, roll back again and on it.**

