Batch vs Online

Tuesday, December 3, 2019 7:21 PM

Note Written by: Zulfadli Zainal

Main focus: Can the data learn on the Fly/ incrementally from incoming stream of data ?

Batch Learning

In batch learning, the system is incapable of learning incrementally. It must be trained using all the available data.

Usually, training will take a lot of time and computing resource - so typically done offline.

Process Flow:

1. System is trained (Offline)
2. Launched to production
3. Runs without learning anymore. Just applied what it just learned. (We called this Offline Learning)

So, if we have some changes in new data set: Need to retrain new model and replace old system!

Luckily, ML training and evaluation process can be automated -> **So even batch learning system can adapt to change.**

Usually training takes long time: So typically just train new data every 24hrs or weekly (Based on situation)

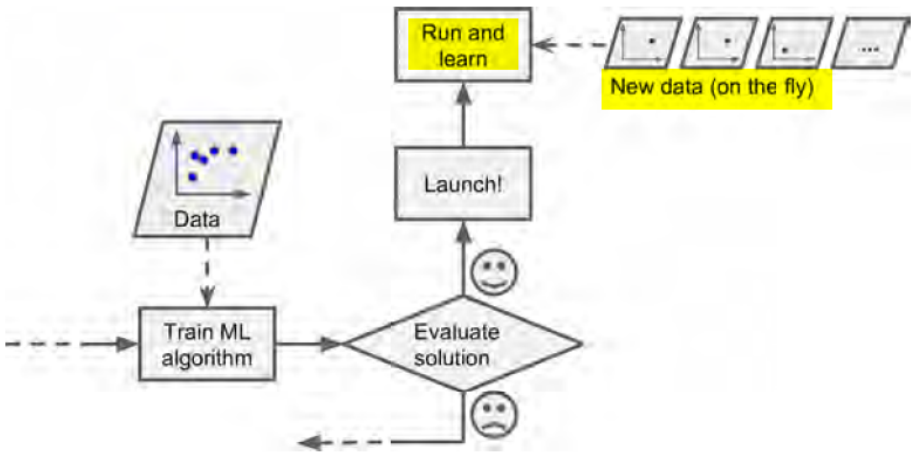
However!

Learning every new data sets from beginning every periodic time might be very difficult if the data sets is very very very big. It might cost you more than expected (CPU resources, and Time).

That's why Online Learning is available.

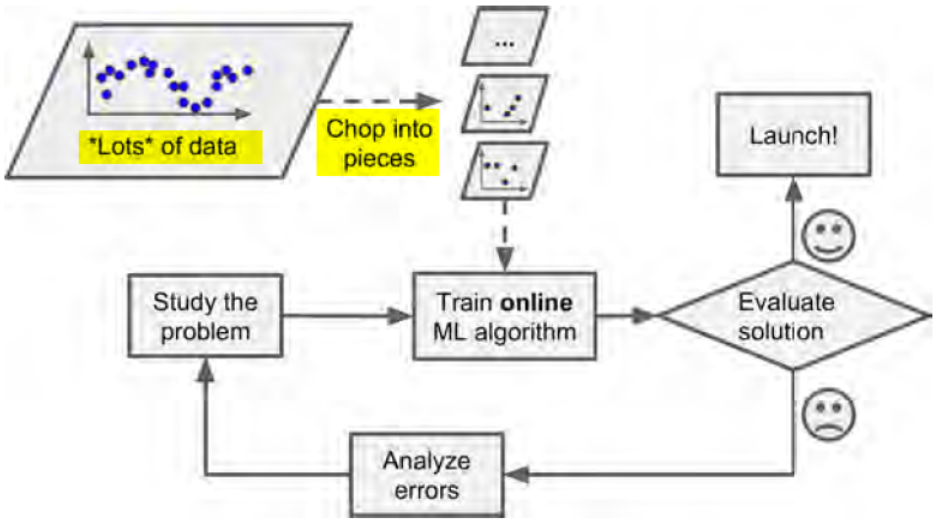
Online Learning

In online learning, you train the system incrementally by feeding it data instances sequentially, either individually or by small groups called mini-batches.



Eg: Stock Prices!

Also can be utilized to train huge sets of data that is out of machine memory capacity.



Important Parameter -> Learning Rate: how fast they should adapt to changing data.

If set HIGH: System will rapidly adapt to new data, but it will also tend to quickly forget the old data.

If set LOW: vice versa.

What is -ve side of Online Learning: Gradually system performance will reduce. Need to closely monitor performance.