Data Science Career Track
Batch Learning vs Online Learning

Batch Learning

An ML algorithm performs Batch Learning if the system can't learn incrementally, and must be trained using all the available data. Since this takes both time and computation power, the learning process typically occurs offline (this is known as 'offline learning'). The system is trained and then launched; it doesn't continue to learn after it has launched. If new data is acquired, a new version of the system needs to be trained to replace the predecessor. Data scientists can automate the training, evaluation, and launch of ML systems that use Batch Learning.

There are shortcomings to an automated Batch approach. Often, you'll need your ML algorithm to respond to changes in vast quantities of data in real-time. It can be quite costly to make a fresh Batch system from this data every day.

Online Learning

Online Learning, on the other hand, trains the system by breaking the data up into small groups and feeding the system those groups over a longer period of time. The learning is broken up into individually cheap and fast steps, which allows for receiving large amounts of data in real-time. Online Learning systems can also adapt quickly, even if

one has limited computational resources. You can set the learning rate (i.e; the speed with which the system adapts to changing data) yourself, but you need to be judicious. If you make this value too high, your system will quickly adapt to the new data at the expense of learning done on previous data; if you set it too low, it will learn about the new data too slowly to be effective.

We must also be careful about the possibility, with Online Learning, that the system's performance is slowed by low-quality data being fed to it unbeknownst to us. Whereas with Batch Learning, you have a relatively high amount of control over the quality of the data your system learns on, with Online Learning, you snooze, you lose (or your model does)!