

Machine Learning in the Cloud

Krish Ganotra

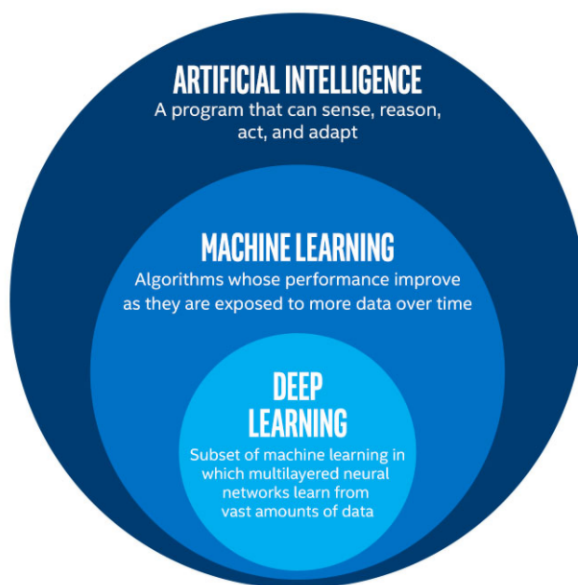
November 2020

1 Introduction

Cloud computing and machine learning are two very powerful, growing technologies that become even more powerful as complements. Machine Learning's few drawbacks are ameliorated with the benefits of cloud computing. We will use Azure to demo how powerful ML in the cloud actually is.

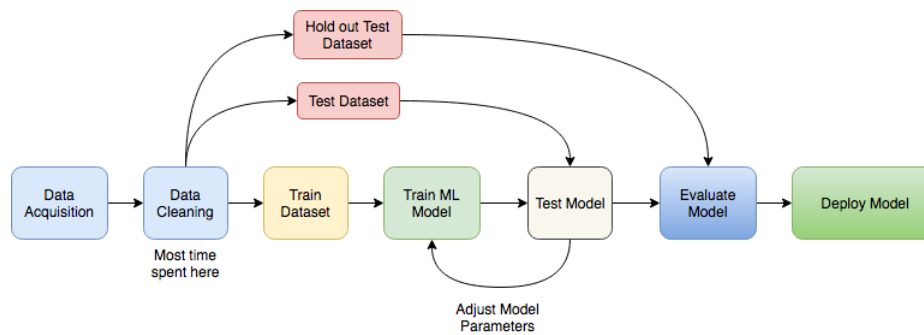
2 Machine Learning

Machine Learning is a rapidly growing field that has become prevalent in all fields with applications everywhere. This powerful tool of making computers do tasks previously thought only humans could do has been able to accomplish extraordinary feats such as detecting cancer in early stages, driver cars, detect fake news, and beat world-class masters in chess.



2.1 Basics

1. First of all, you need data to train your model on. This is generally labeled data (supervised) such as an image with a label for what the image is of, or it can be non-labelled data (unsupervised) such as many images that need to be categorized by the objects in them.
2. Clean, purify, and format the data
3. Split your data into 80% training data, 20% testing data. Your model will use the training data to "learn" the patterns of the data and then will be tested against the testing data to measure how good of a model it is.
4. Choose a model (e.g. neural network or one of its many variations, SVMs, Decision Trees etc.) and choose optimal hyperparameters. Hyperparameters are variables that the programmer needs to manually define that dictates how the model will train.
5. Train your model on the training data and evaluate the model using the testing data.
6. Deploy your model so it can be used in real-time!



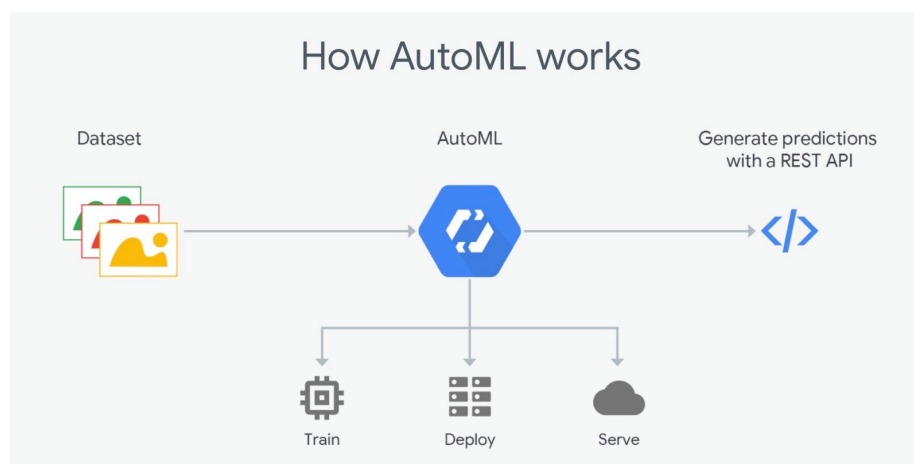
2.2 Drawbacks

- Lots and lots of data – hard to store and access
- Extremely computationally expensive to train models
- Extensive knowledge is required to train models by coding them straight-up and still requires lots of time to achieve a useful model accuracy
- Setting up ML environments with all the necessary packages is difficult to install and maintain

hint, these problems can be solved with cloud computing :)

3 Benefits of ML in Cloud

1. **Compute Power** - As mentioned earlier, Machine learning is a very computationally expensive task. The resources needed to train models quickly and effectively are only readily accessible over the cloud. These resources are also easily scalable in the cloud making your machine learning tasks as efficient and affordable as possible.
2. **Storage Resources** - Machine learning also requires immense amounts of data (the more the data the better!) which is more easily stored and accessed in the cloud than on local machines.
3. **AutoML** - With cloud computing, anyone can train and test ML models with ease even with minimal AI or ML knowledge. Experts in machine learning are constantly working on making ML more easily accessible for everyone, creating AutoML services and deploying them in the cloud for us to use.
4. **ML Environment** - all the packages needed for ML come pre-installed with ML instances. You don't have to worry about installing and maintaining these packages since the cloud provider does this for you.



4 Public Cloud Provider Comparison

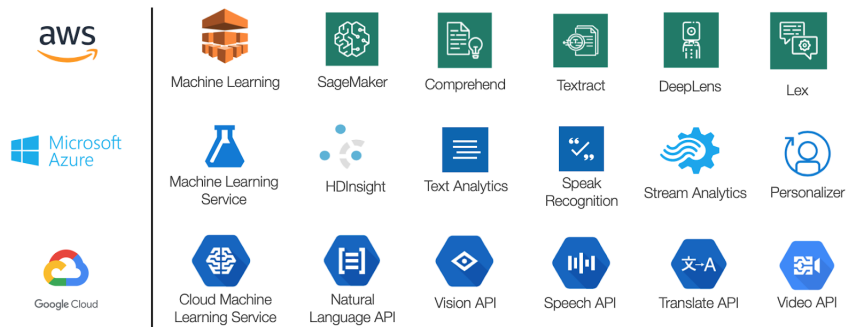
AWS, GCP, and Azure all provide services for speech, vision, natural language processing, video intelligence, and AutoML. AWS and GCP additionally offer services for translation and conversational interface (e.g. alexa).

4.1 AutoML

- AWS Sagemaker Studio

- Azure Machine Learning Studio
- GCP AutoML

Machine Learning Services From Leading Cloud Providers



5 Demo

A few ML Qwiklabs hands-on labs are linked below with descriptions. Each of these cost 1 credit. Unfortunately, none of them use AutoML which is one of the most powerful tools provided by these public cloud providers. This is because many of the AutoML labs on Qwiklabs cost more than 1 credit. Feel free to explore AutoML on your own time (and funding).

- https://www.qwiklabs.com/focuses/581?catalog_rank=%7B%22rank%22%3A13%2C%22num_filters%22%3A1%2C%22has_search%22%3Atrue%7D&parent=catalog&search_id=7715435 This is an introduction into GCP's AI Platform. In this lab, you create a new ML instance, use a jupyter notebook, clone ML code from github, and train a ML model to predict on housing data
- https://www.qwiklabs.com/focuses/7639?catalog_rank=%7B%22rank%22%3A13%2C%22num_filters%22%3A1%2C%22has_search%22%3Afalse%7D&parent=catalog In this lab, you create a normal VM and learn how to set it up to run machine learning code. This is probably the most broadly applicable lab since it will help you in learning how to run code on a VM. This is a powerful tool that is used universally in the IT field. For this lab, the tensorflow code and data is given to you and all you must do is set the environment up to run this code on the data.
- https://www.qwiklabs.com/focuses/2157?catalog_rank=%7B%22rank%22%3A11%2C%22num_filters%22%3A1%2C%22has_search%22%3Atrue%7D&parent=catalog&search_id=7857263 In this lab you use BigQuery Machine Learning. This is an extremely powerful tool if **you are familiar**

with SQL. It allows you to train a machine learning model with a simple SQL query on data stored in GCP's BigQuery