AutoML Primer

Created by Shravanthi Sridhar
Last updated Apr 29, 2020 by Valmeek Kudesia • 3 min read • Analytics

What is AutoML?

Automated Machine Learning provides methods and processes that make Machine Learning available for non-Machine Learning experts. Individuals can use autoML to get started in the use of and accelerate research in Machine Learning. AutoML can also be viewed as automated methods for model selection and/or hyper-parameter optimization. In many cases of ML use, beginners feel as if they are guessing as they test out different hyper-parameters for a model. AutoML provides a guided way to select models and optimize hyper-parameters. It can also be useful to get a sense of what model performance is possible for a problem.

There are several automl libraries for this purpose e.g. autoweka, auto-sklearn, h20, tpot (see below for references). Regardless of library or model chosen, data must be cleaned prior to invoking automated methods.

Types of API's available:

Auto-sklearn

- Built upon sklearn
- Good for small or medium sized data
- Includes feature engineering and thus easier for beginners
- Creates pipeline & optimizes via Bayesian search
  - includes **meta-learning** to initialize Bayesian optimizer
  - **automated ensemble construction** from configurations evaluated during optimization

Tree Based Pipeline Optimization Tool (TPOT)

- Implements genetic programming for optimization
- Is an extension of sklearn with its own regressor and classifier methods
- Automated evaluation of thousands of possibilities for winning approach
- Important caution: TPOT has weakness with categorical strings. These must be integer coded prior to use of TPOP

H20

- Supports both R and python platform.
- Has visualization

Autokeras:

- Based upon keras deep learning framework
- API design is similar with scikit; which simplifies its usage

- Useful if deep learning models are required in a problem.; relatively few problems at this time require deep learning models. Most problems are approachable by basic machine learning algorithms with configuration or tweaks.
- Automates
  - Search for architecture and hyper-parameters for deep learning models
  - NAS (neural architecture search) to design neural nets

Cloud AutoML

- Simple user-friendly UI
- Does not require coding knowledge to deploy models and derive insights
- Leverages neural architecture search (NAS) and Transfer Learning to address complex problems like object segmentation, etc. Of note, NAS is often very computationally expensive.
- Typically, **reinforcement learning** or **evolutionary algorithms** are used to design the new neural net architectures
  - Allows for discovery of architectures far more complicated than manually designed
  - May be optimized for particular goals.
- From schematic below:
  - 2 blocks: model selection and Deep Learning NAS. Most problems are approachable by basic machine learning models with configuration or tweaks
  - When deep-learning models are deployed, the framework deploys DL-NAS instead of basic models

The Bottom Line: What can we do with this?

Many teams want to evaluate the potential value if machine learning models are used with data in different scenarios. AutoML is a good starting point to try various models, evaluate different results, and derive insights from them. Many teams at CCA work with data and they may use AutoML to leverage data and extract patterns. This tool also provides an easier way to express results to a business audience.

For people without a coding background, CloudML with its simple UI is the preferred place to start and evaluate methods. For people more familiar with ML, AutoML/CloudML is quick way to identify likely successful ML models and establish baseline performance and then state-of-art performance could be achieved via manual tuning. Data cleaning and a good feature matrix is a absolute prerequisite.

Any form of Automl provides insights regarding patterns in data. However, in many cases the results cannot be used as the end result. Domain knowledge SMEs must still verify if the selected features make sense or are reasonable.

AutoML is less helpful for individuals who want to learn the details of machine learning as the individual processes are not shown. However, AutoML may serve as an easy entry to cultivate interest.

Do check these out to get started:

Google Cloud AutoML: https://cloud.google.com/automl/docs

Scikit-learn API: https://scikit-learn.org/stable/modules/classes.html

Auto-sklearn API: https://automl.github.io/auto-sklearn/master/api.html

TPOT API: https://epistasislab.github.io/tpot/api/

AutoKeras API: https://autokeras.com/tutorial/overview/

H2o API: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/welcome.html

Like Be the first to like this

No labels