AUTOML

An Overview of Automated Machine Learning

DR. SRIJITH RAJAMOHAN

CONTENT OUTLINE

What is AutoML?

Hyperparameter optimization

Bayesian optimization

Neural Architecture Search

Auto-Sklearn example

Auto-Keras example

WHAT IS AUTOML?

- Machine Learning is inherently an iterative process.
- A lot of time spent selecting the right learning technique/model family and then fine-tuning the right hyperparameters
- AutoML is a suite of tools and techniques.
- Minimize the amount of time spent optimizing both hyperparameters and architectures.

WHERE DO I LOOK?

- Visit AutoML.org at https://www.ml4aad.org/automl/ for updated information on the state-of-the-art.
- The define it as:

”Automated Machine Learning provides methods and processes to make Machine Learning available for non-Machine Learning experts, to improve efficiency of Machine Learning and to accelerate research on Machine Learning.”

WHAT IS AUTOMATED?

- As complexity of problems evolve, off-the-shelf packages become popular
- Not quite AI: Humans still need to be in the loop
- However, a lot of the things can be automated such as model selection, hyperparameter optimization and validation.
- Goal: Minimize the need for ML experts in the model development cycle, Humans still need to critically analyze the results

WHAT IS AUTOMATED?

- What Machine Learning algorithm to use?
- How to preprocess the data?
- How to set the hyperparameters? E.g. step size in gradient descent, tree depth in decision trees, regularization coefficient etc.
- Auto-WEKA calls it the Combined Algorithm Selection and Hyperparameter optimization

MODEL SELECTION TOOLS

- Auto-sklearn builds upon the functionality of scikit-learn
- Hyperopt uses Bayesian optimization whereas TPOT uses genetic programming
- H2O AutoML does model selection
- Auto-Keras performs architecture search, allows you to automate the generation of models in Keras

HYPERPARAMTER SEARCH

- Grid Search: Systematically try every combination of hyperparameters on a grid, rigorous and computationally expensive
- Random Search: Sample randomly from your search space, known to work as well or better than grid search
- Evolutionary algorithms
- Bayesian optimization: Use historical information to refine the search space, shown to be quite efficient

BAYESIAN OPTIMIZATION

MATH ALERT

- Bayesian optimization uses probabilistic modeling to estimate the relationship between hyperparameter space and the model’s performance
- Computes P(score|configuration) and uses ’Acquisition functions’ such as Expected Improvement, Probability of Improvement, bandit-based criterion etc.
- Iteratively updates the model

BAYESIAN OPTIMIZATION

MATH ALERT

- Classfied based on a probabilistic regression ’Surrogate function’ choice
- Gaussian Process, Random Forests, Tree-structured Parzen Estimators
- Response surface used to approximate ML model is gradually built iteratively from the history
- Common to use the Expected Improvement criterion ’which is high in regions of low predictive mean and high predictive variance’

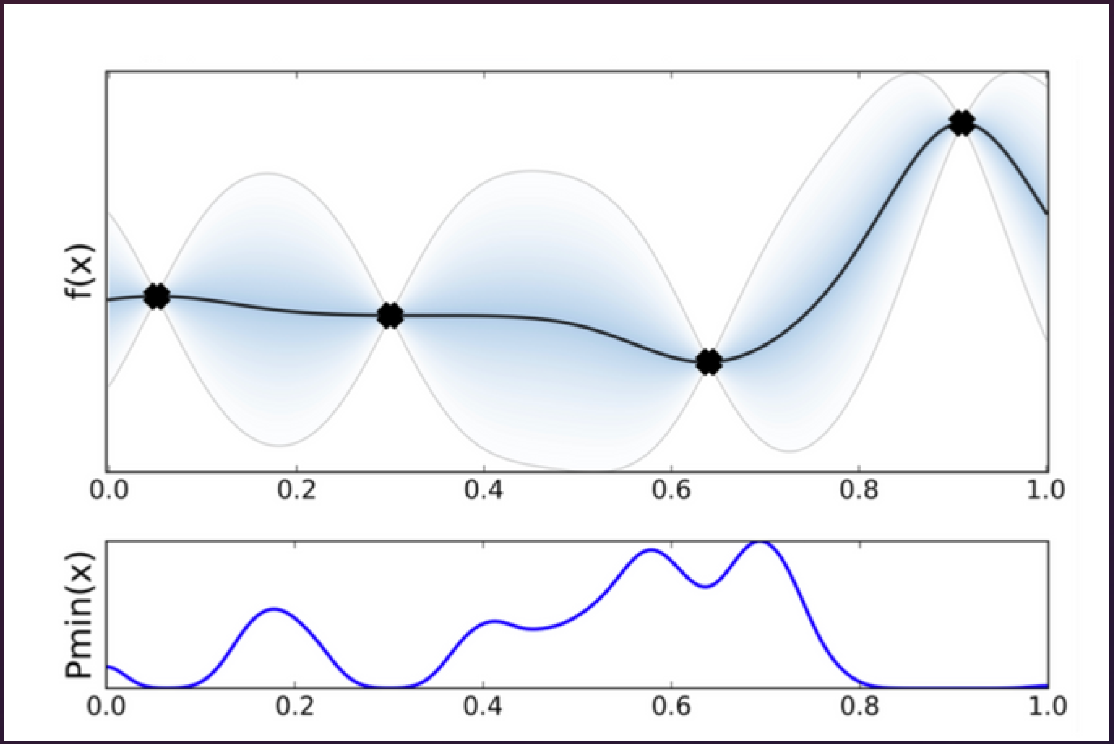
EXPECTED IMPROVEMENT

MATH ALERT

Acquisition function

$$EI(x) = \int_{fbest}^{\infty} (f - f_{best})p(f|x)df$$

Image from Cedric Archchambeau' slides
Image courtesy of Javier Gonzalez



SOME BAYESIAN OPTIMIZATION TOOLS

- Gaussian Process-based approaches work best in low-dimensional space while Tree-based models work best in high-dimensional space
- For Tree-based approaches, you have Random-Forest-based SMAC and Tree-structured Parzen estimators (TPE)
- Hyperopt uses TPE
- Auto-sklearn uses SMAC while Auto-Keras is a Neural Architecture Search tool that uses Bayesian optimization

THE SMAC ALGORITHM

- Sequential Model-based Algorithm Selection (SMAC) uses Random Forests or approximate Gaussian Process to determine viable configurations
- Based off the Sequential model-based optimization (SMBO) algorithm but can also handle categorical variables with Hamming distance function kernel
- Uses the Expected Improvement criterion
- The next point, to evaluate, is selected that has the highest EI criterion

META-LEARNING

- Used in Auto-sklearn
- Uses characteristics or attributes of the data, called meta features, that can be computed efficiently
- They can be number of points, features, classes etc.
- Used to determine what algorithm to use on unseen data
- Provides a warm-start for Bayesian optimization

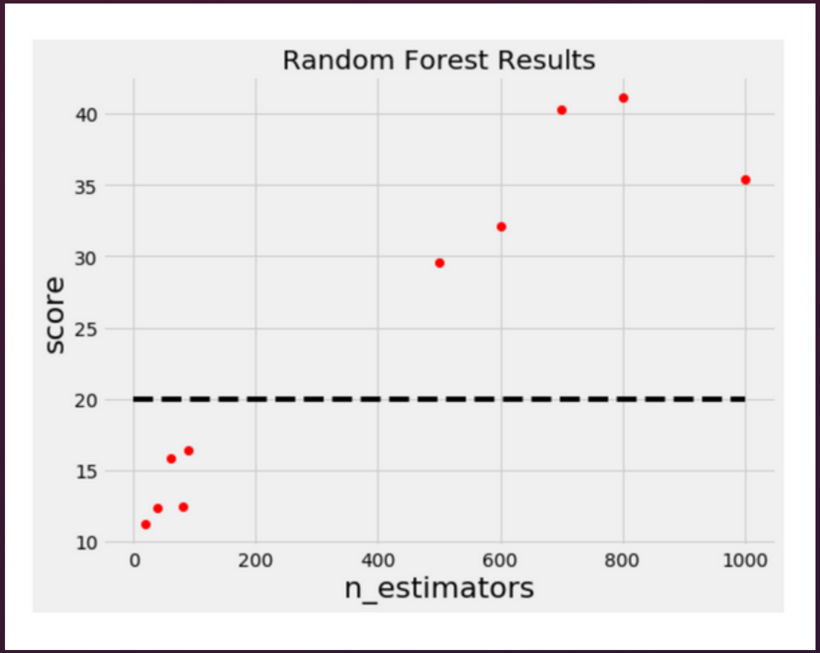
TREE STRUCTURED PARZEN ESTIMATORS

MATH ALERT

- Uses $p(\text{configuration}|\text{score})$ instead of $p(\text{score}|\text{configuration})$
- Uses densities as a surrogate function
- Configuration selected by maximizing EI which is proportional to the ratio of $l(x)/g(x)$, the ratio of the density lower than a threshold to that greater than the threshold
- Limitation: cannot optimize hyperparameters independently of each other

TREE STRUCTURED PARZEN ESTIMATORS

MATH ALERT



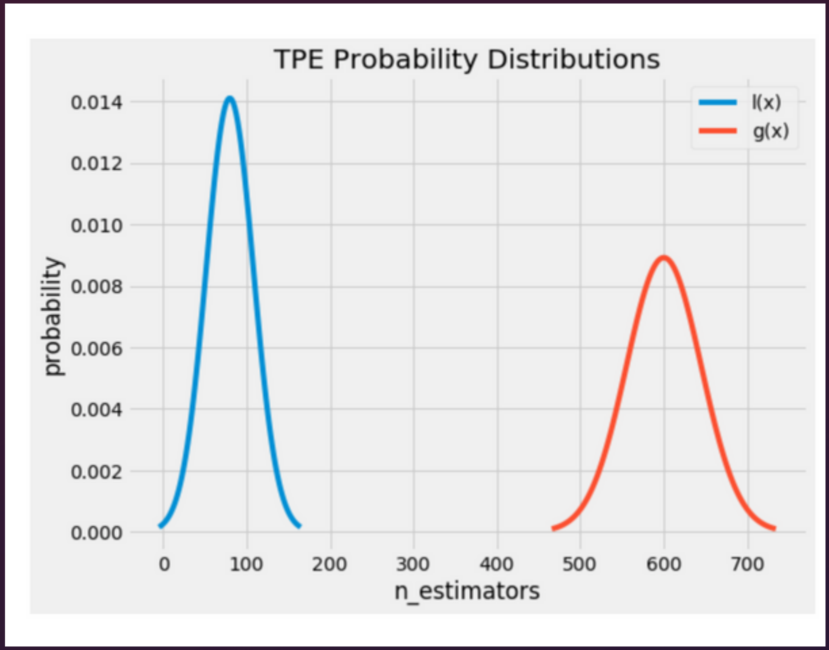


Image from https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machine-learning-b8172278050f

NEURAL ARCHITECTURE SEARCH

- Most Neural Network architectures created by people, a very time consuming iterative process
- Neural Network ’architecture engineering’ can be automated by NAS
- Adjust the the number of layers, nodes and the connectivity through an optimization process
- Three key components: search space, search strategy and performance estimation strategy

NEURAL ARCHITECTURE SEARCH

- Very time and resource intensive process
- Deep architectures for Computer Vision problems almost impossible to do without a cluster of GPUs
- Google Cloud AutoML: an end-to-end solution for doing AutoML on the Google Cloud provides this option

AUTO-KERAS

- Auto-Keras still a very nascent tool
- Uses Network morphism to train Neural Networks, no training from scratch
- Ability to explore smaller architectures
- Propose approximate edit-distance and kernels to morph networks in a scalable manner
- Hybrid approach for optimization that uses a hybrid tree-structured search/simulated annealing approach