Automated Machine Learning (Auto-ML)
Gisele Lobo Pappa
Postgraduate only
This course will be taught in English
Machine Learning course desirible but not required.

When faced with a new machine learning (ML) task, selecting which algorithm to use to effectively perform the task at hand is considered a very hard problem. The area of Automated Machine Learning (Auto-ML) emerged to address this problem by using search or optimization methods to determine the best ML algorithm (and its respective hyper-parameters) for a given target dataset from a set of candidate algorithms.

Objective: The main objective of this course is to introduce Auto-ML, presenting its main concepts, categorization, search techniques and applications. Being a very active area of research, part of the course will be based on seminars considering recent papers on the area.

Syllabus:

- Introduction to Auto-ML:
    - Motivation: Why do we need Auto-ML techniques?
- Historical Perspective:
    - Machine learning versus Optimization
    - Auto-ML versus Meta-Learning
- Auto-ML's categorization:
    - Automatic configuration of ML algorithms / Automatic setting of hyperparameters
    - Automatic selection and configuration of ML algorithms
    - Automatic selection or construction of ML pipelines
- Review of ML
    - Preprocessing techniques
    - Classifiers
    - Ensembles
- Search Techniques:
    - Bayesian Optimization-based methods
    - Bandit-based methods
    - Evolutionary-based methods
    - Hybrid methods and others
- Auto-ML frameworks:
    - Auto-WEKA
    - Auto-sklearn
    - Tree-Based Pipeline Optimization (TPOT)

- REsilient ClassifIcation Pipeline Evolution (RECIPE)
- Hyperband

- Main Applications
  - Based on research papers from the top machine learning/data mining conferences

Bibliography:

- AutoML: Methods, Systems and Challenges. Frank Hutter, Lars Kotthoff, Joaquin Vanschoren (Eds), https://www.automl.org/book/
- D, Golovin, B. Solnik, S. Moitra, G. Kochanski, J. Karro, and D. Sculley. Google vizier: A service for black-box optimization. In Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), ACM, 1487-1495, 2017.
- A. G. C. de Sá, W. J. G. S. Pinto, L. O. V. B. Oliveira, G. L. Pappa. RECIPE: A grammar-based framework for automatically evolving classification pipelines. In Proc. of the European Conference on Genetic Programming (EuroGP), 246-261, 2017.
- L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar. Hyperband: Bandit-based configuration evaluation for hyperparameter optimization. In Proc. of the International Conference on Learning Representations (ICLR), 2017.
- R. S. Olson, N. Bartley, R. J. Urbanowicz, and J. H. Moore. Evaluation of a tree-based pipeline optimization tool for automating data science. In Proc. of the Genetic and Evolutionary Computation Conference (GECCO), 485-492, 2016.
- M. Feurer, A. Klein, K. Eggensperger, J. Tobias Springenberg, M. Blum, and F. Hutter. Efficient and robust automated machine learning. In Proc. of the International Conference on Neural Information Processing Systems (NIPS), 2755-2763, 2015.
- G. L. Pappa, G. Ochoa, M. R. Hyde, A. A. Freitas, J. Woodward, and J. Swan. Contrasting meta-learning and hyper-heuristic research: the role of evolutionary algorithms. Genetic Programming and Evolvable Machines 15 (1), 3-35, 2014.
- C. Thornton, F. Hutter, H. H. Hoos, and K. Leyton-Brown. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), ACM, 847-855, 2013.