

Aprendizado de Máquina

Aula 6.1 - Auto ML


Adriano Rivolli

rivolli@utfpr.edu.br

Especialização em Inteligência Artificial

Universidade Tecnológica Federal do Paraná (UTFPR)
Câmpus Cornélio Procopio
Departamento de Computação

Conteúdo

- 
- 1 AutoML
 - 2 Auto-sklearn

AutoML

Introdução

- Cada algoritmo/técnica tem um viés
- Especialista escolhe a melhor solução
 - ▶ Experiência
 - ▶ Tentativa e erro
 - ▶ Força bruta
- Automatização da escolha

Vantagens

- Apoio ao especialista e usuário leigo
- Reduz atividades repetitivas e/ou monótonas
- Simplifica o processo
- Gera bons resultados
- Permite ampla utilização (desburocratiza)

Ponta a ponta

- Entrada: conjunto de dados
- Saída: solução de AM
- Etapas:
 - ▶ Exploração/interpretação dos dados
 - ▶ Engenharia de dados: pré-processamento
 - ▶ Indução do modelo: escolha do algoritmo e hiperparâmetros
 - ▶ Pós-processamento: solução esperada

O que pode ser automatizado?

- Os **processos** que serão utilizados (*pipeline*)
- Os **algoritmos** que serão utilizados em cada processo *pipeline*
- Os **hiperparâmetros** dos algoritmos utilizados em cada processo do *pipeline*

Componentes chaves

- Espaço de busca
 - ▶ Envolve todas as escolhas (*pipeline*, algoritmos, hiperparâmetros)
- Estratégia de busca
 - ▶ Como encontrar as soluções dentro do espaço de busca
- Avaliação do desempenho
 - ▶ Como avaliar o desempenho das soluções candidatas

Tópicos adicionais

■ Otimização

- ▶ Define um espaço de busca de soluções e uma estratégia de busca

■ Meta-aprendizado

- ▶ Aprendizado de Máquina do Aprendizado de Máquina

■ Busca de arquiteturas de Redes Neurais Artificiais

- ▶ Descobertas de novas arquiteturas de Redes Neurais Profundas

Principais abordagens

- Seleção de algoritmo/modelo
- Otimização de hiperparametros (HPO)
- *Combined Algorithm Selection and Hyperparameter optimization (CASH)*

Ferramentas de Auto ML

- Auto-WEKA
- **Auto-sklearn**
- H2O AutoML
- TPOT
- Soluções comerciais (Google, Amazon, Oracle, Microsoft)

Auto-sklearn

Auto-sklearn

■ Versão 1.0 lançada em 2015

- ▶ 5 algoritmos de classificação
- ▶ 14 métodos de tratamento de atributos
- ▶ 4 métodos de pré-processamento

■ Versão 2.0 lançada em 2020

- ▶ Classificação, regressão, *multi-target*
- ▶ Restrição de tempo
- ▶ Customização
- ▶ Explicação dos modelos

Visão geral

■ Limite de recursos

- ▶ Memória e tempo
- ▶ Paralelização

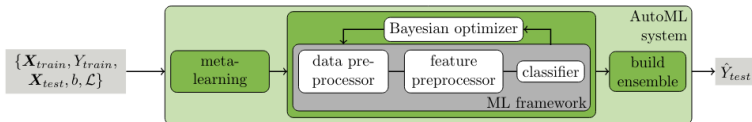
■ Espaço de busca

- ▶ É possível restringir o espaço de busca
- ▶ Desativar etapas do *pipeline*
- ▶ Especificar um subgrupo de algoritmos

■ Seleção do modelo

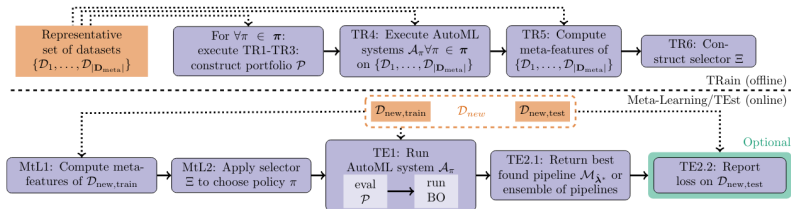
- ▶ Diferentes estratégias para obtenção do modelo final
- ▶ Usa comitê das melhores estratégias

AutoML 1.0



Fonte: Feurer, M., et. al., 2016. Efficient and Robust Automated Machine Learning. Advances in Neural Information Processing Systems.

AutoML 2.0



Fonte: Feurer, M., et. al., 2020. Auto-Sklearn 2.0: Hands-free AutoML via Meta-Learning. arXiv:2007.04074. <https://arxiv.org/pdf/2007.04074>