AutoML与推荐系统

AutoML and Recommender Systems

猎聘大数据研究院 单艺

议题

1

PART ONE

缘起

2

PART TWO

超参数优化

3

PART THREE

自动化建模

4

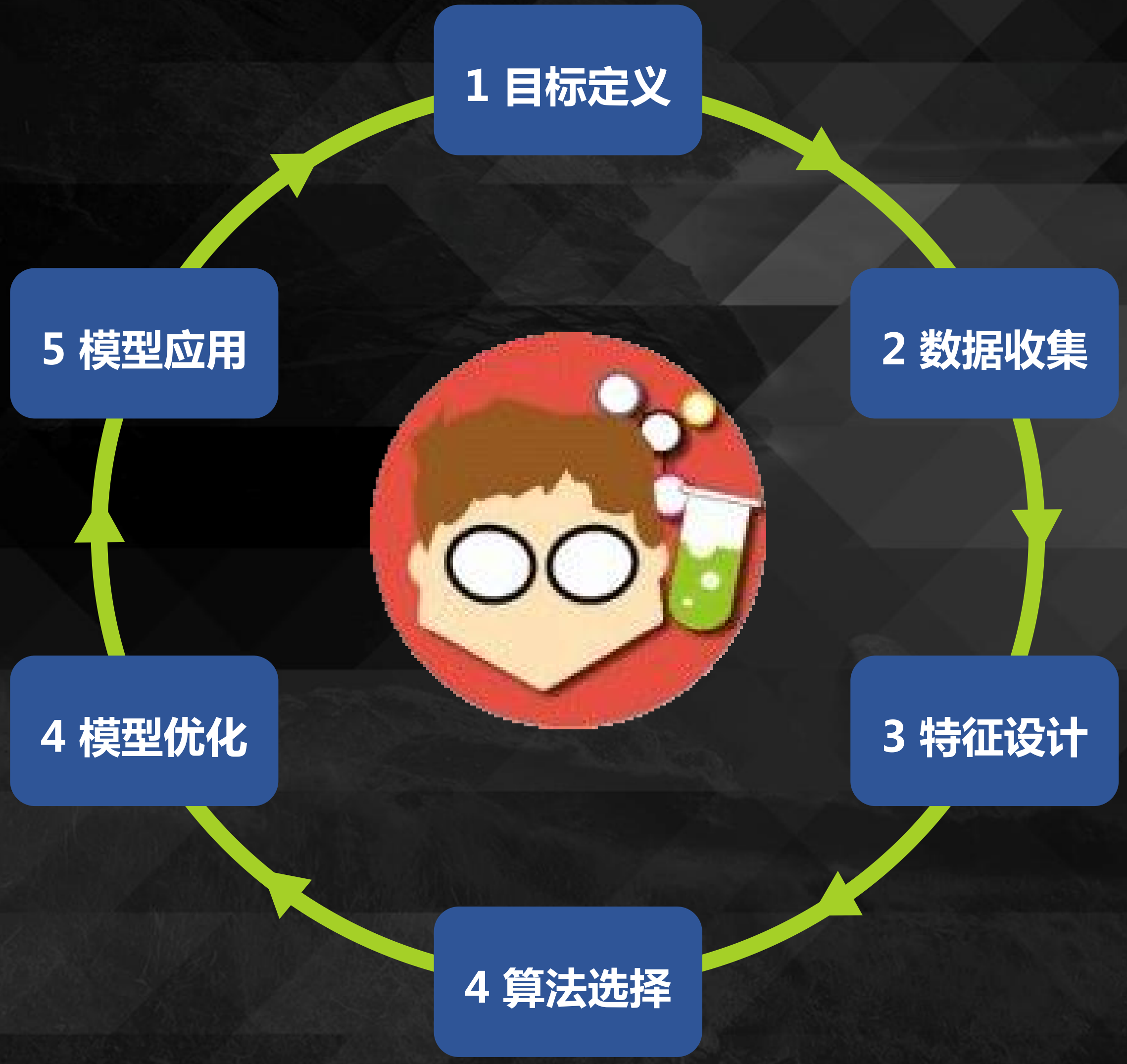
PART FOUR

试验和展望

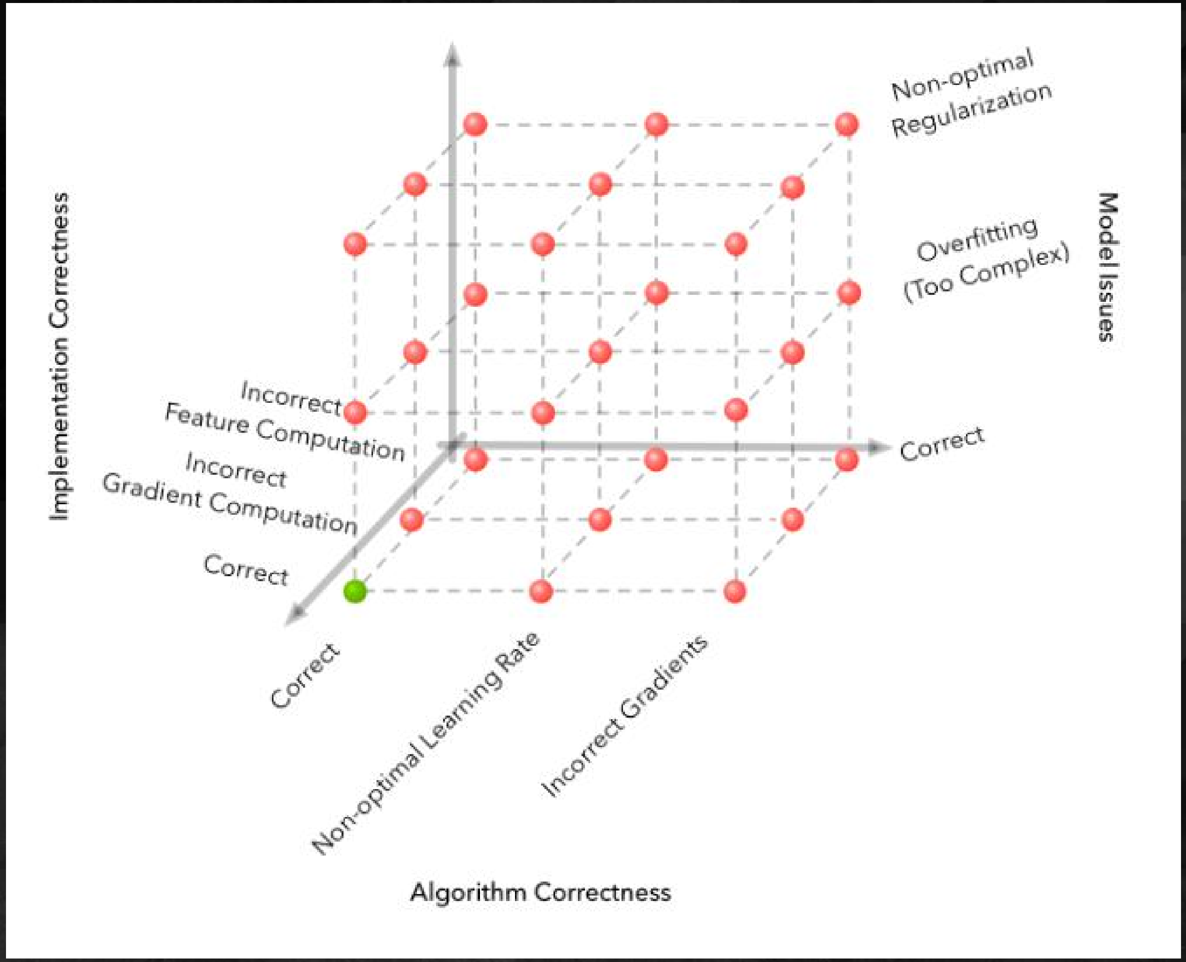
PART ONE

缘起

预测建模过程



Modeling Is Hard



做一个数据科学家是什么体验？

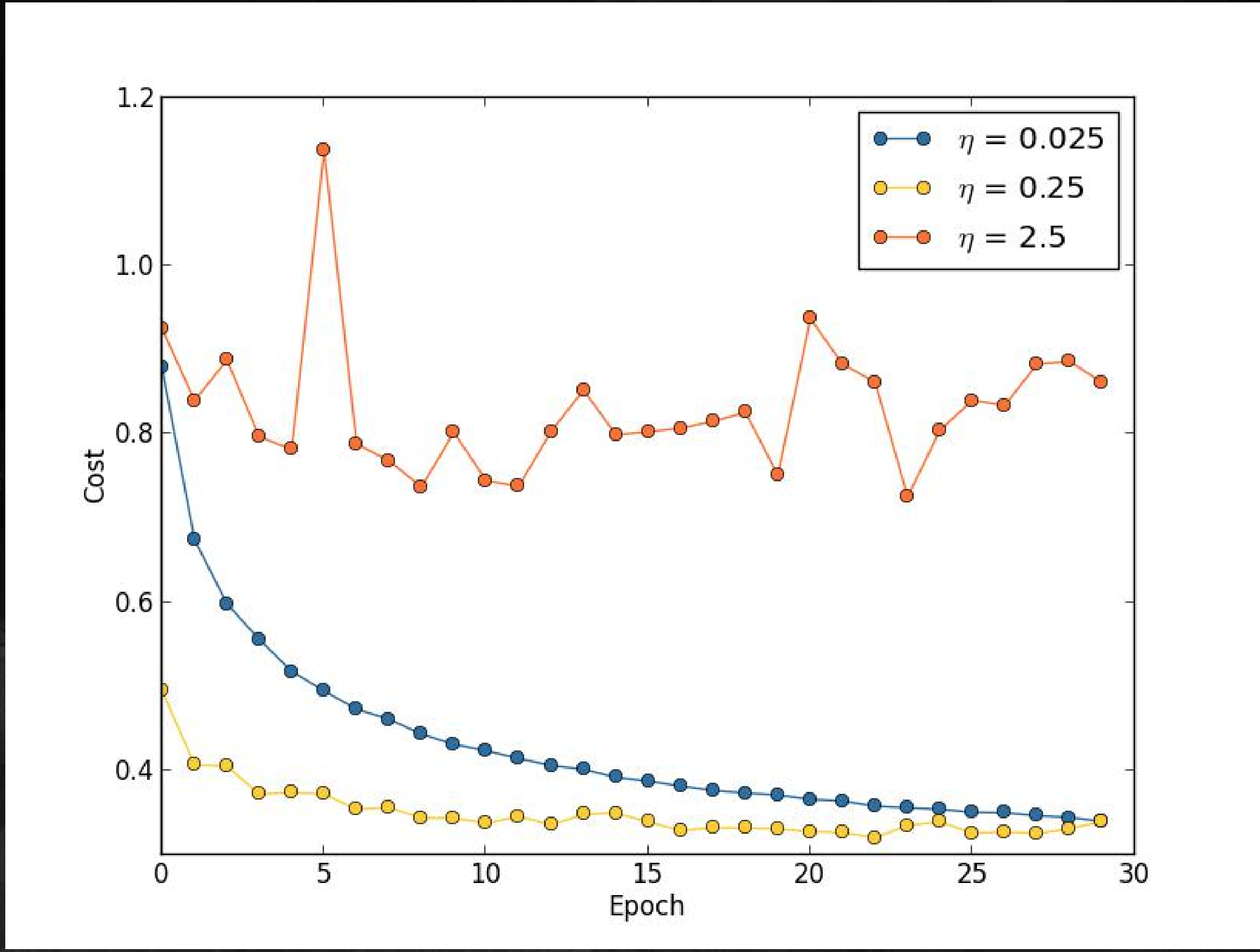
PART TWO

超参数优化

模型/算法超参数

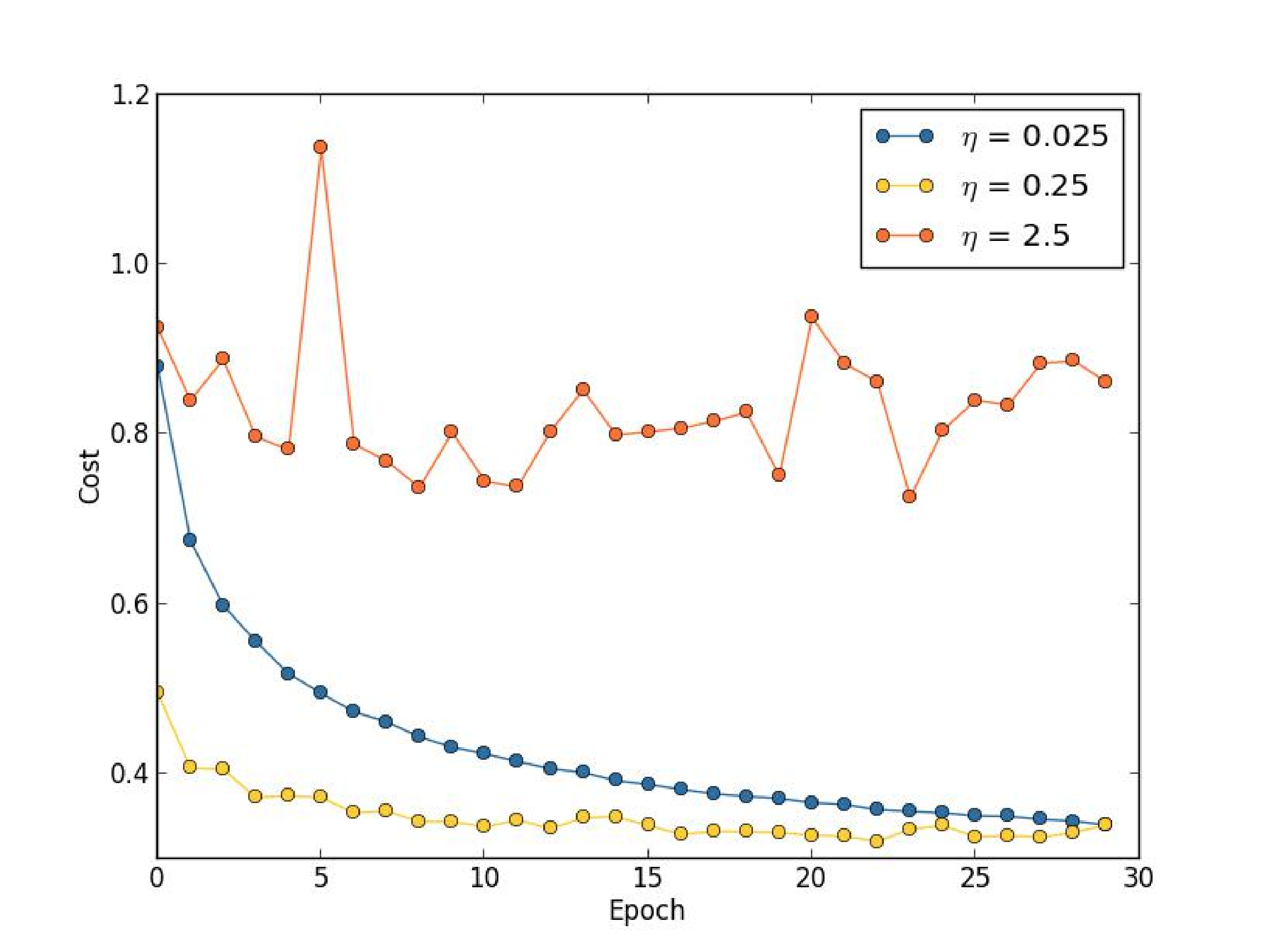
Item CF	相似度算法、相似度因子权重…
Matrix Factorization	隐因子数、正则化权重…
Neural Networks	层数、每层神经元数、dropout比例
GBDT	提升次数、树的最大深度、学习率、样本采样率、特征采样率…
Random Forest	树的数量、树的最大深度、样本采样率、特征采样率…
Logistic Regression	正则化权重、正则化方法
Gradient Descent	学习率、批次大小、迭代次数…

超参数的影响



Neural Networks with Different Learning Rates on MINST

超参数的影响



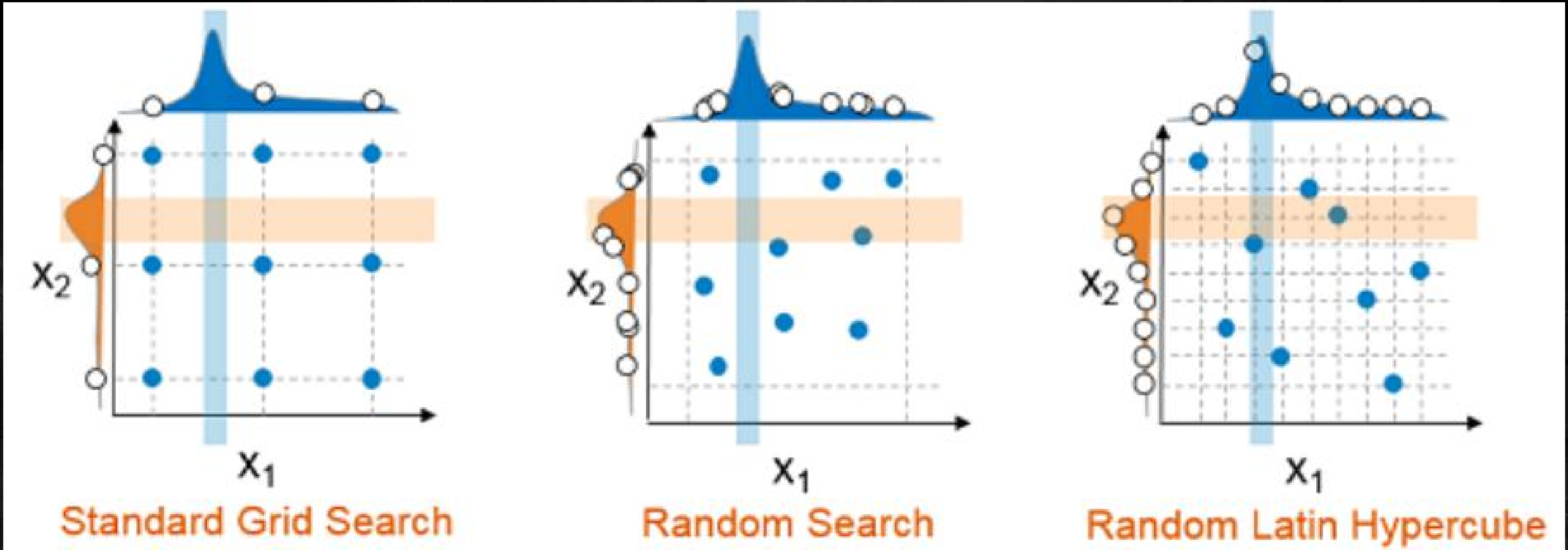
Neural Networks with Different Learning Rates on MINST

超参数优化问题

- 目标：找到在验证数据集上效果最好的超参数
- 挑战：
  - 参数空间巨大
  - 效用函数是一个黑盒子
  - 训练和评估成本高
- 问题：
  - 如何聪明地搜索最佳超参数

手工调参

寻找最佳超参数



x_2
x_1
Standard Grid Search
x_2
x_1
Random Search
x_2
x_1
Random Latin Hypercube

自动超参数优化的主要方法

- 贝叶斯优化：
  - 高斯过程回归
  - SMAC
  - TPE
- 谱模型
- Bandit算法
  - Hyperband算法

贝叶斯优化

1. 假设目标函数符合某个先验分布
2. 初始随机试验
3. 根据观测结果得到后验分布
4. 利用后验分布选取下一个试验点
    - 使用获取函数决定新的试验点

高斯过程回归

A **Gaussian process** is a collection of random variables, any subset of which is jointly normally distributed.

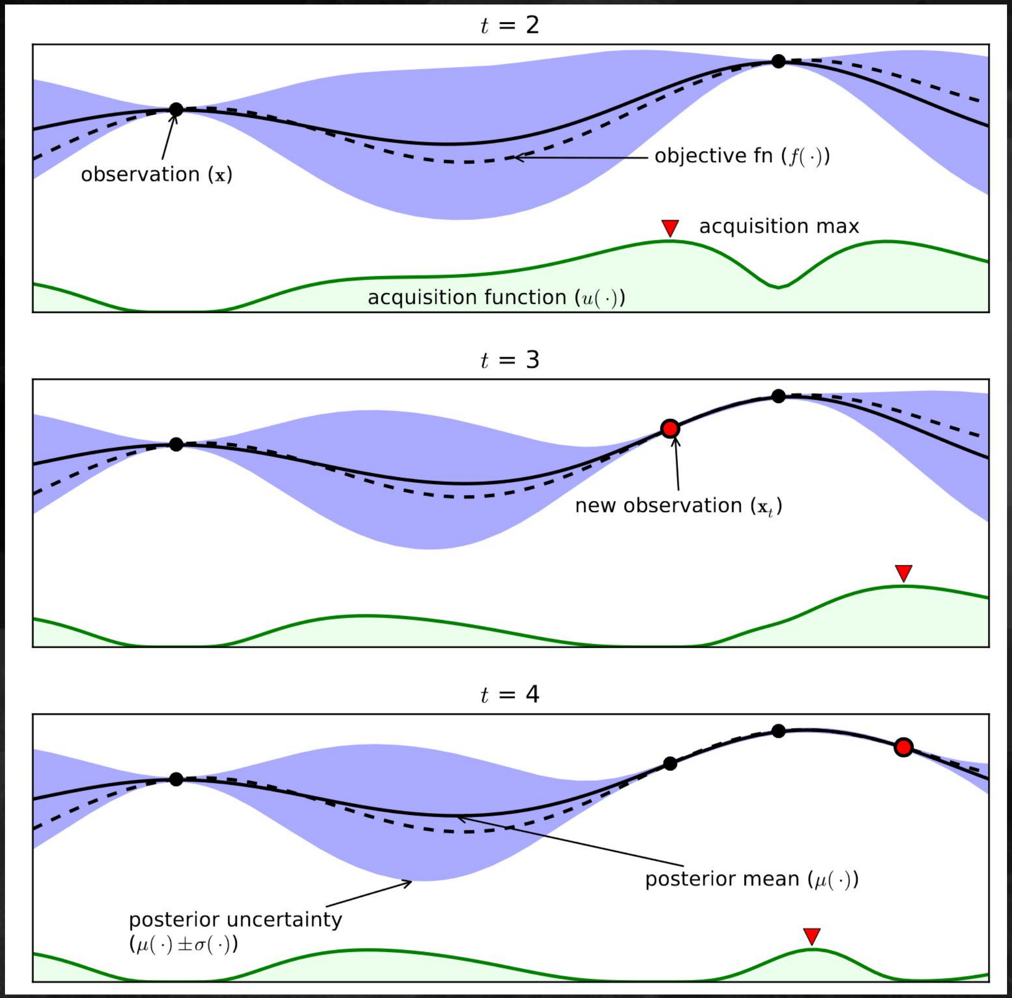
Gaussian process regression:

assume form of mean and covariance among data → functional form

$$y \sim N\left(\mu_y, K\right) \rightarrow y^* \mid y \sim N\left(\mu_y^*, \Sigma^*\right)$$

高斯过程回归

用GPR优化超参数



t = 2
observation ($\mathbf{x}$)
objective fn ($f(\cdot)$)
acquisition max
acquisition function ($u(\cdot)$)
t = 3
new observation ($\mathbf{x}_t$)
t = 4
posterior mean ($\mu(\cdot)$)
posterior uncertainty
($\mu(\cdot) \pm \sigma(\cdot)$)

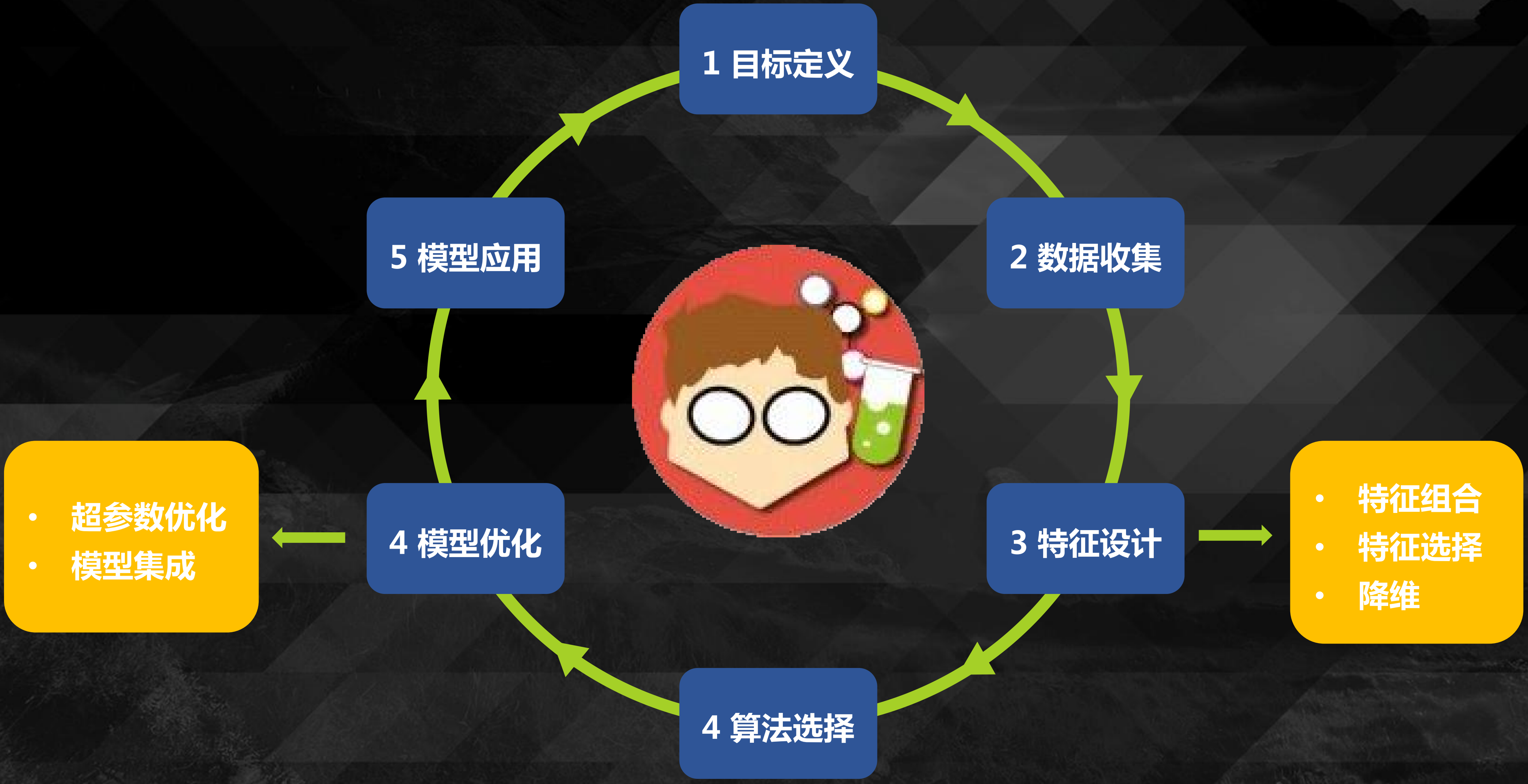
贝叶斯优化软件包

1. Spearmint
2. Yelp MOE -> SigOpt
3. Hyperopt
4. Scikit-optimize
5. SMAC

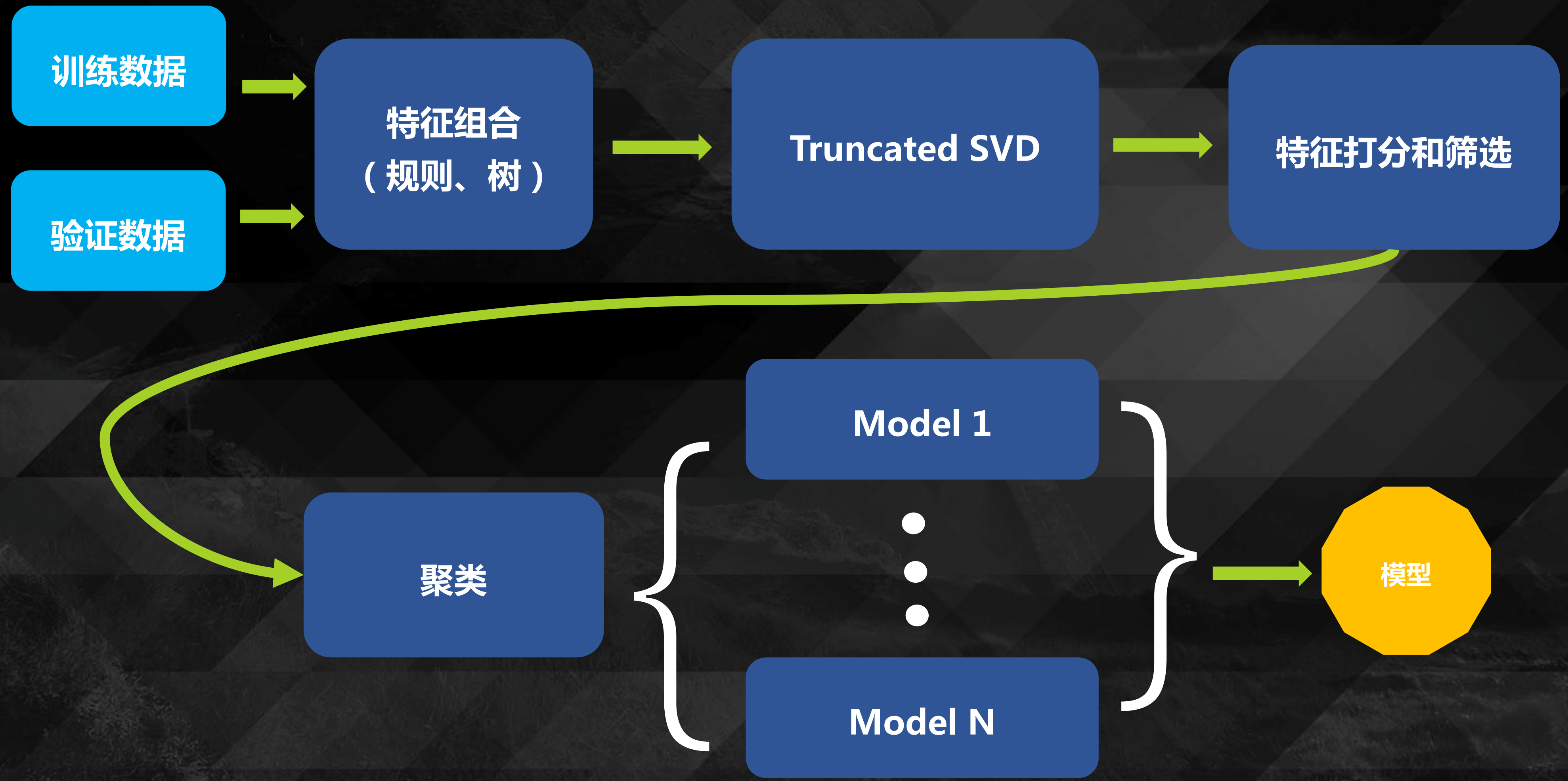
PART THREE

自动化预测建模

预测建模流程



通用自动化预测建模系统



Network Architecture Search

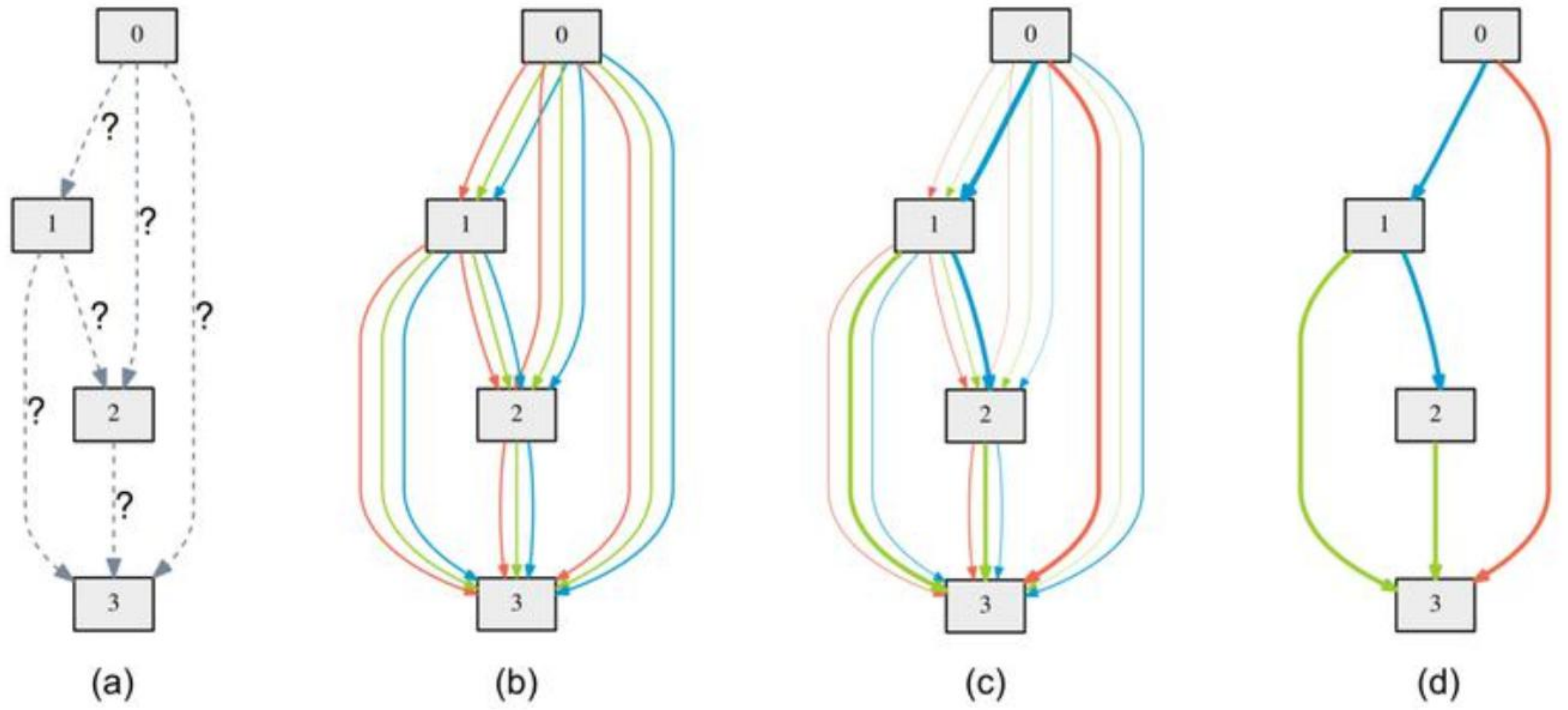
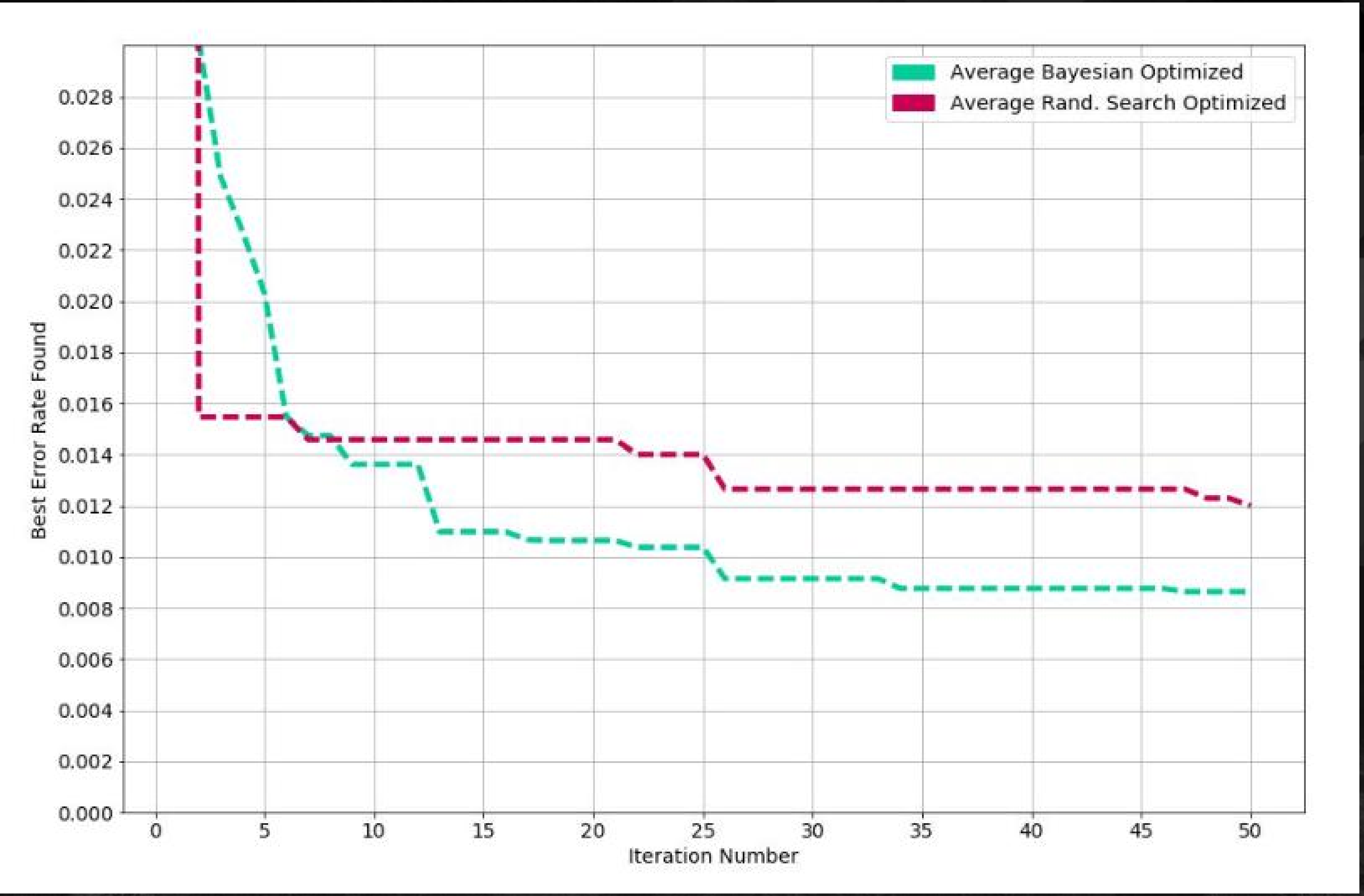


Figure 1: An overview of DARTS: (a) Operations on the edges are initially unknown. (b) Continuous relaxation of the search space by placing a mixture of candidate operations on each edge. (c) Joint optimization of the mixing probabilities and the network weights by solving a bilevel optimization problem. (d) Inducing the final architecture from the learned mixing probabilities.

PART FOUR

试验和展望

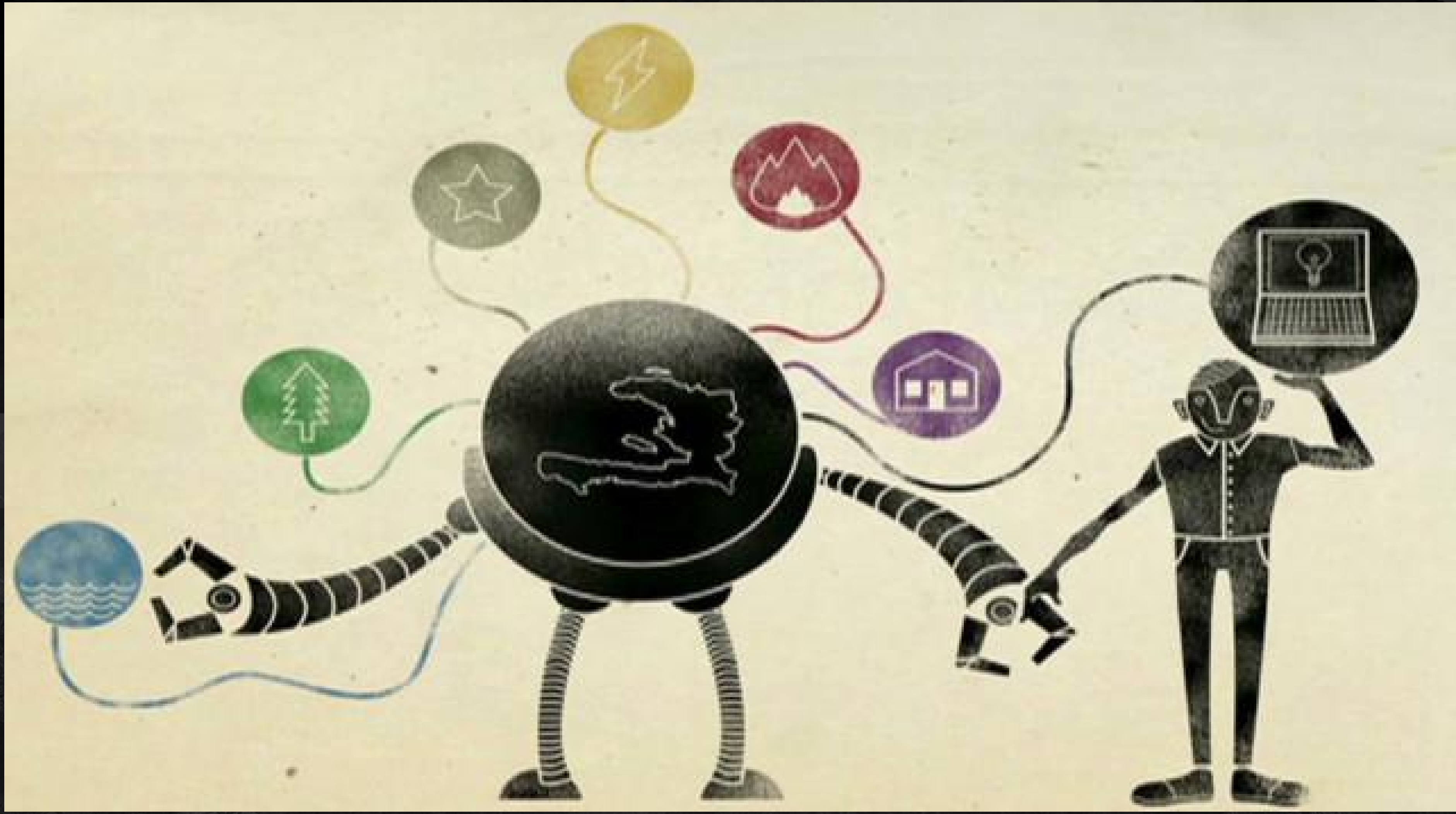
试验对比



Average Bayesian Optimized
Average Rand. Search Optimized
Best Error Rate Found
Iteration Number
0.028
0.026
0.024
0.022
0.020
0.018
0.016
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0
5
10
15
20
25
30
35
40
45
50

算法吃人？

人机协作



THANK YOU

猎聘大数据研究院