自动化预测建模初探

Automatic Predictive Modeling: A Bayesian Approach

猎聘大数据研究院 单艺

- 现任猎聘首席数据官，兼职就业顾问
- 曾任职于美国Altera、Yahoo、奥美
- 经验：数据挖掘、搜索、广告、招聘
- 兴趣：数据挖掘和商业分析
- 毕业于清华大学和美国亚利桑那大学

议题

1 PART ONE 缘起

2 PART TWO 超参数优化

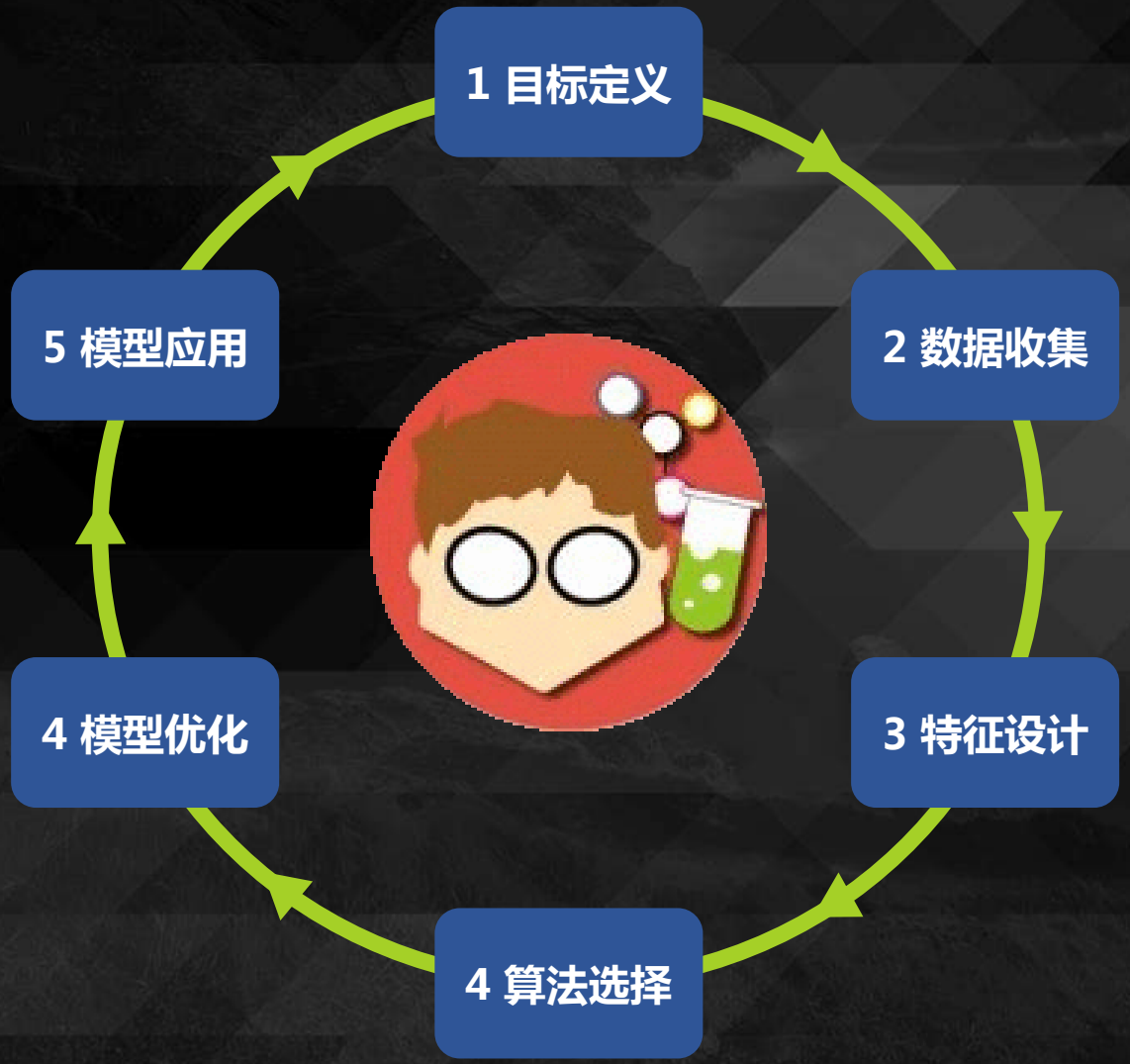
3 PART THREE 自动化建模

4 PART FOUR 试验和展望

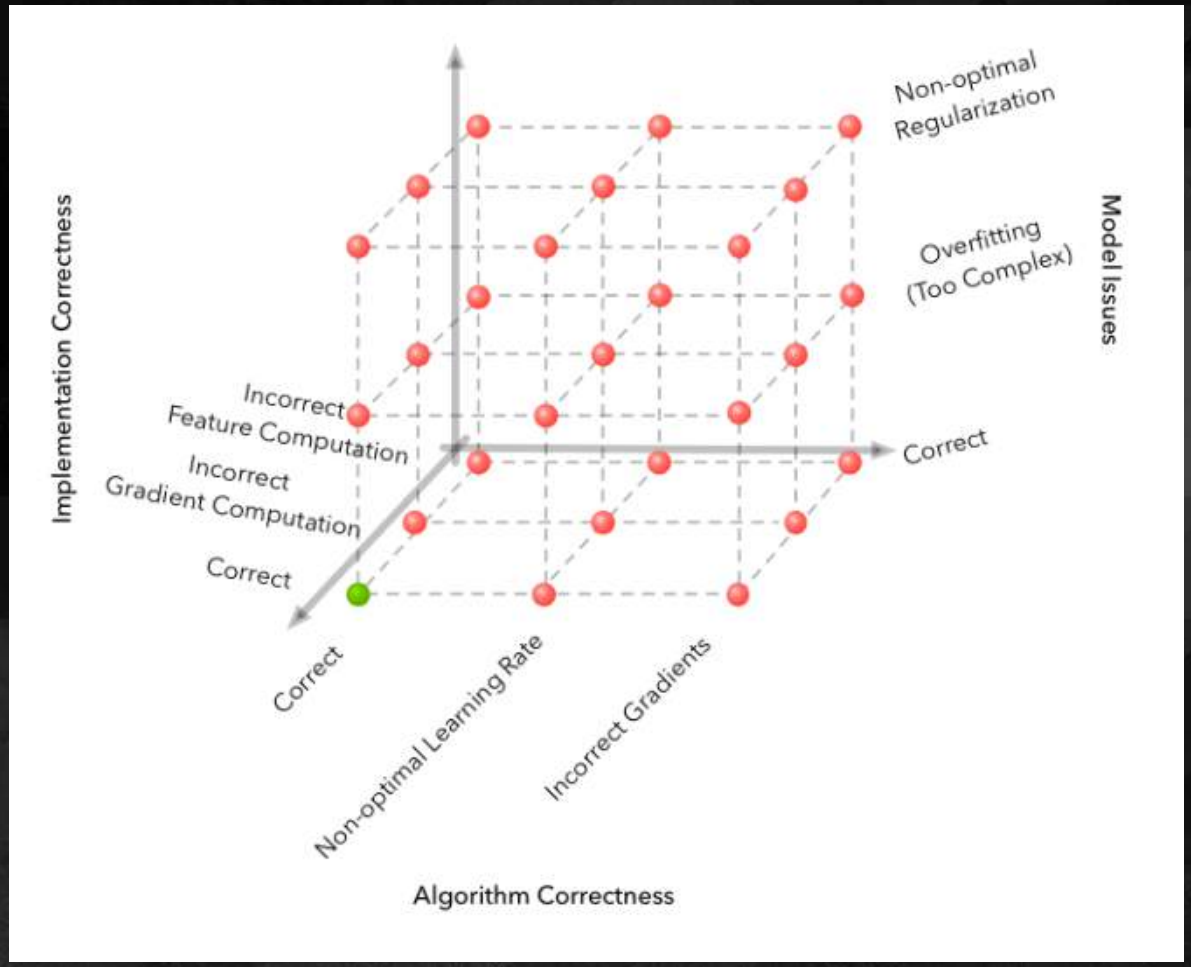
PART ONE

缘起

预测建模过程



Why Is Modeling So Hard?



做一个数据科学家是什么体验？

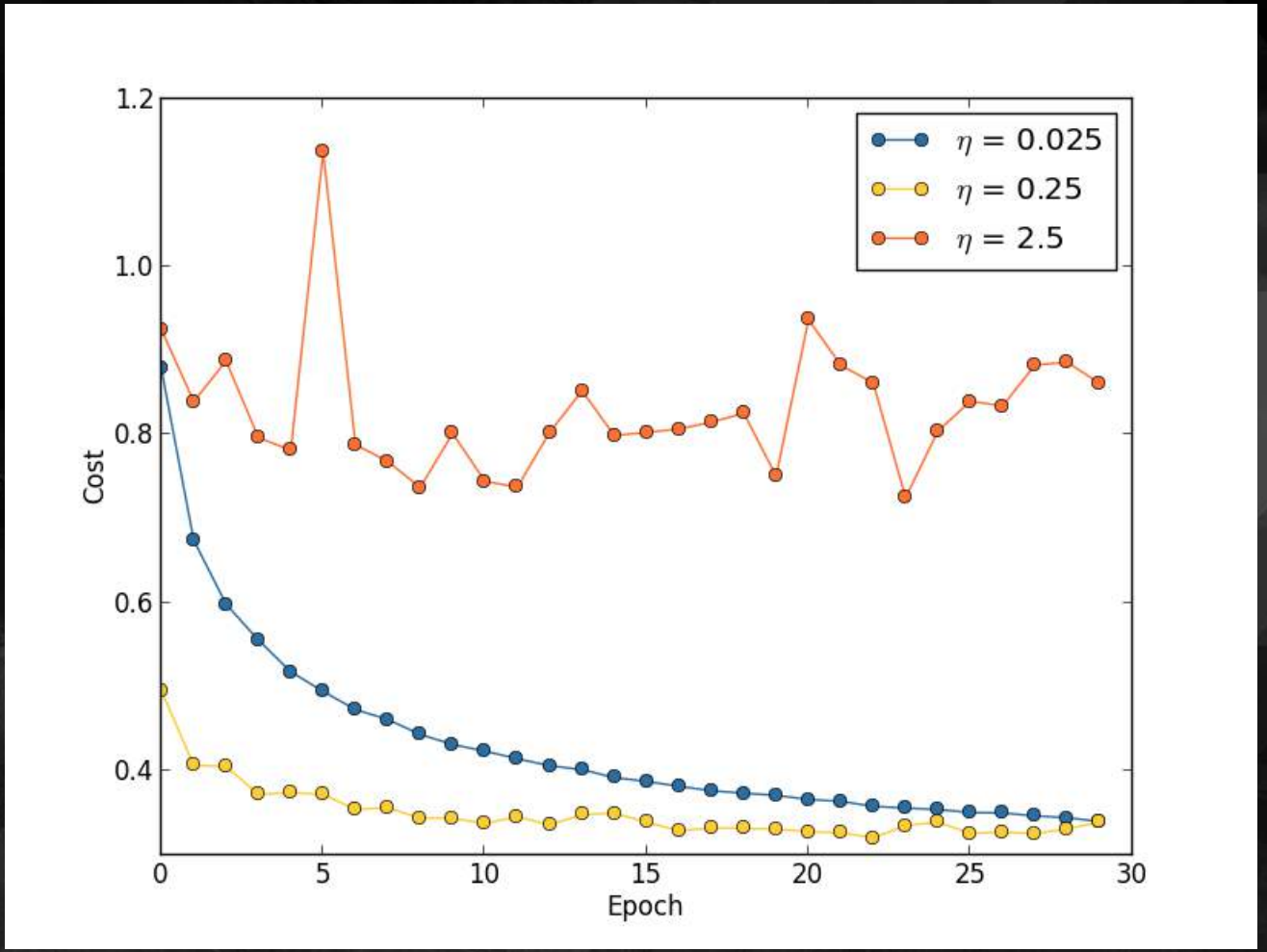
PART TWO

超参数优化

模型/算法超参数

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
SVM	惩罚参数、核参数、不敏感参数ε
LDA	主题数量、先验分布参数（α、β）
Gradient Descent	学习率、批次大小、迭代次数…

超参数的影响



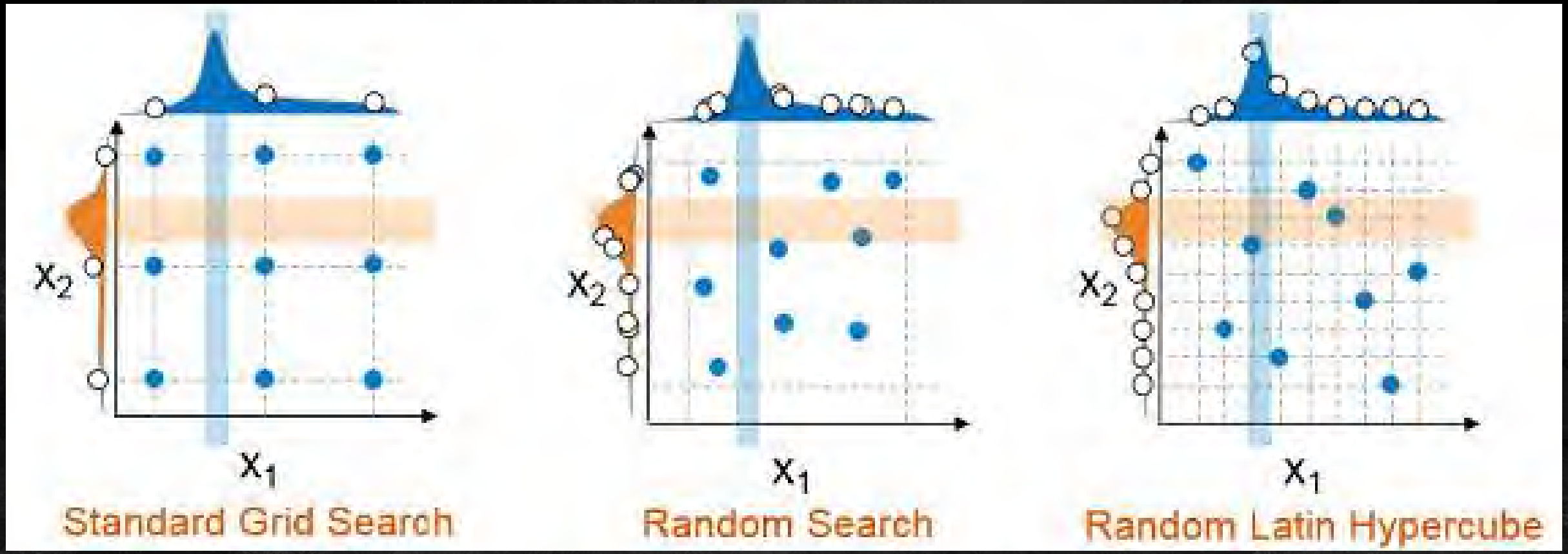
Neural Networks with Different Learning Rates on MINST

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

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

- 通过获取函数（acquisition function）决定新的试验点

效用概率模型：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

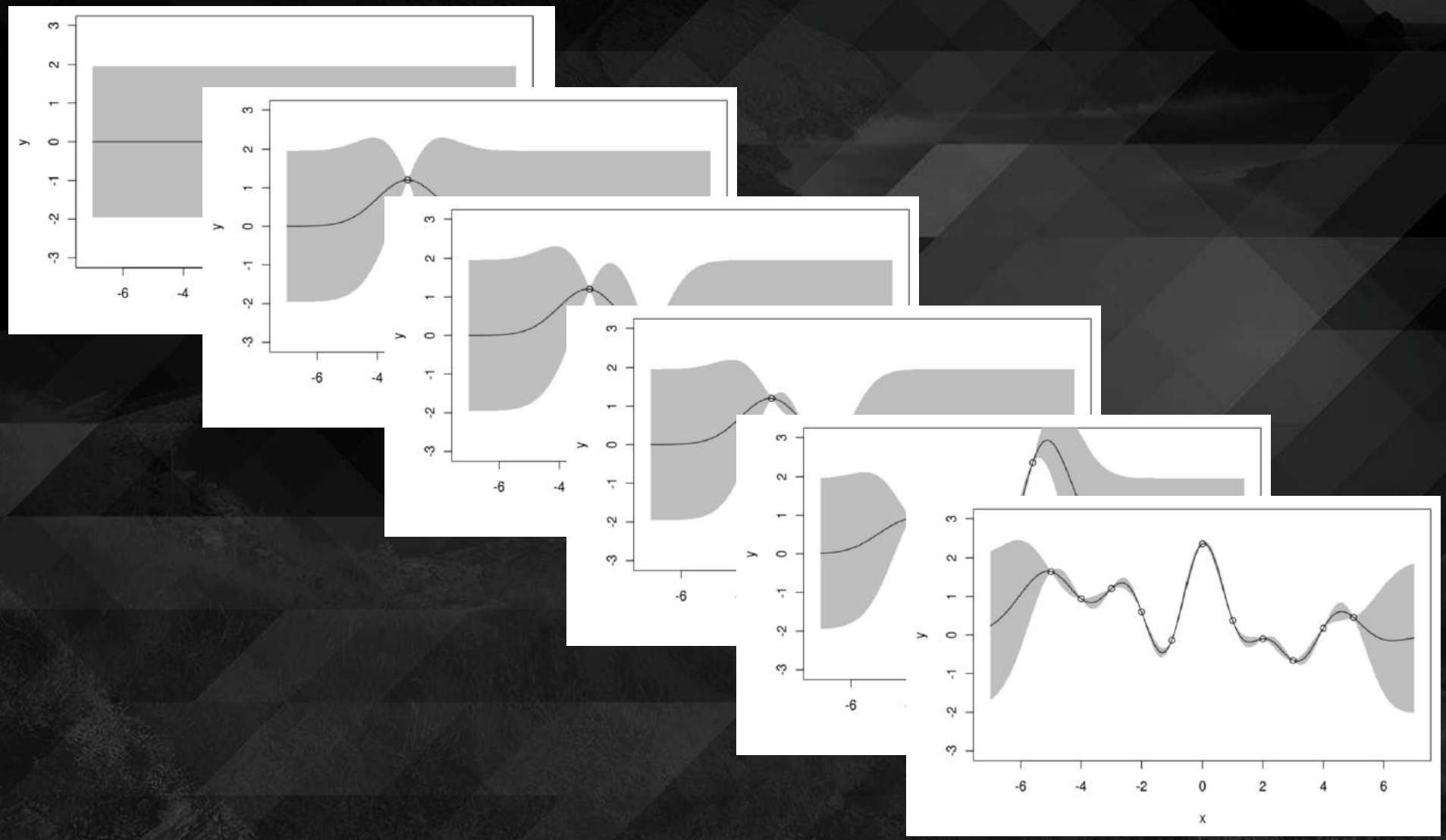
$$p(y^*|x^*,\mathbf{x},\mathbf{y}) \sim \mathcal{N}(\mu,\sigma^2)$$

$$\mu = k(x^*,\mathbf{x})^T(k(\mathbf{x},\mathbf{x})+\sigma^2_{noise}I)^{-1}\mathbf{y}$$

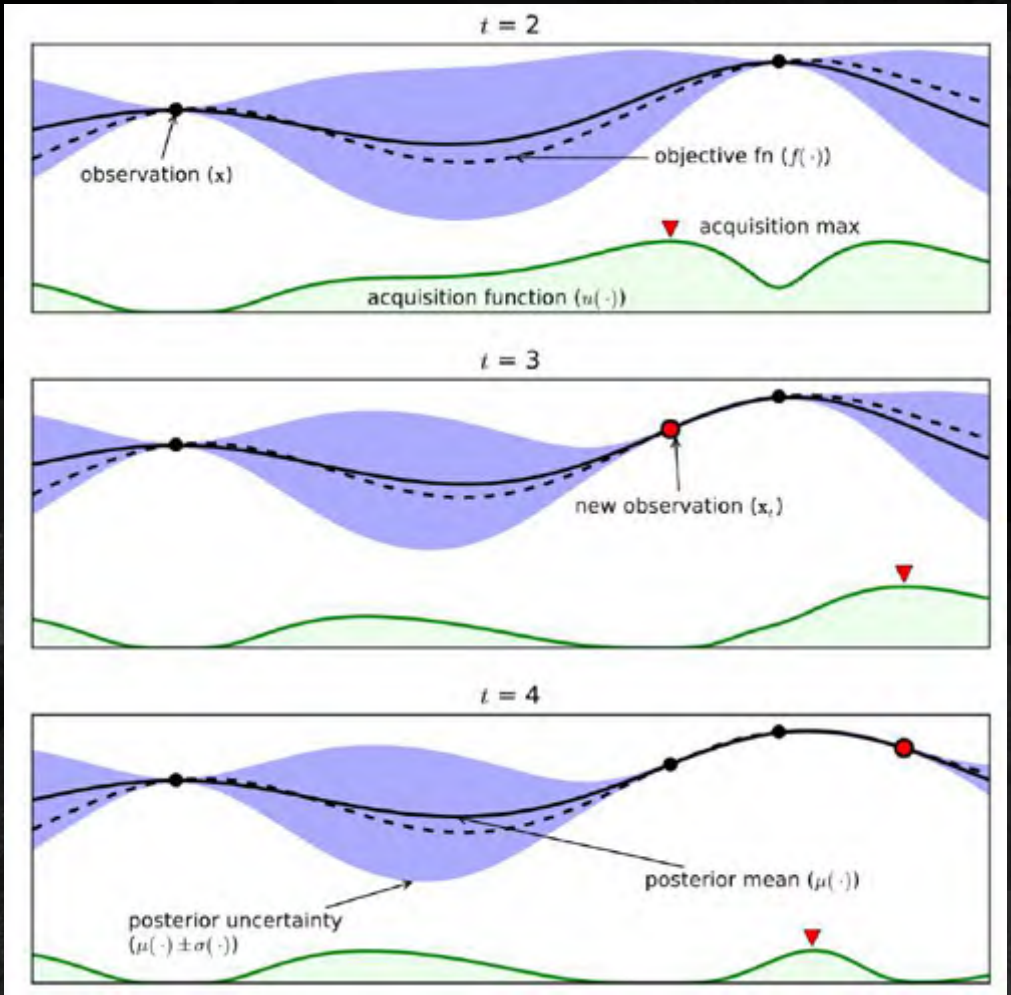
$$\sigma^2 = k(x^*,x^*) - k(x^*,\mathbf{x})^T(k(\mathbf{x},\mathbf{x})+\sigma^2_{noise}I)^{-1}k(x^*,\mathbf{x})$$

$$k(x_1,x_2) = \exp\left(-\frac{\| x_1 - x_2 \|^2}{2l}\right)$$

高斯过程回归



用GPR优化超参数



Acquisition Functions:

- Probability of Improvement
- Expected Improvement
- Upper Confidence Bound

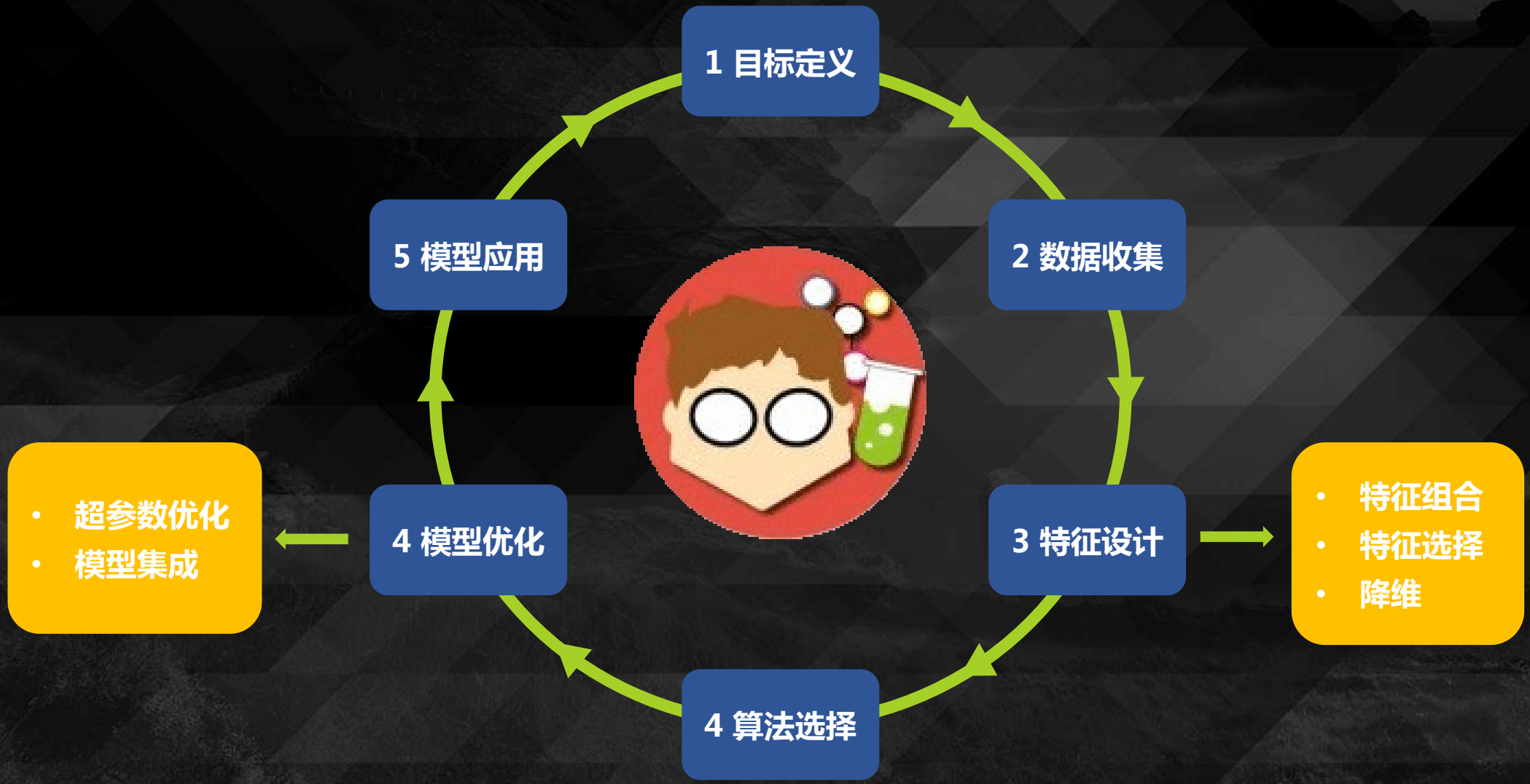
优化软件包

1. Spearmint
2. Yelp MOE -> SigOpt
3. Hyperopt
4. Scikit-optimize
5. SMAC
6. 其他：近似梯度方法

PART THREE

自动化预测建模

预测建模流程



PART FOUR

试验和展望

算法吃人？

人机协作

谢谢聆听