知识

对machine learning相关知识和理论的储备。

XGBoost

原理？

决策树节点分裂时是如何选择特征的？

Gini Index和Information Gain的公式，并举例说明？

回归树和分类树的区别？

与Random Forest作比较，并以此介绍什么是模型的Bias和Variance？

XGBoost的参数调优有哪些经验？

XGBoost的正则化是如何实现的？

XGBoost的并行化部分是如何实现的？

梯度提升树GBDT

原理？

优点：不需要做特征的归一化；自动进行特征选择；模型可解释性较好；可以适应多种损失函数，如SquareLoss、LogLoss等。

缺点：boosting是个串行的过程，不能并行化；计算复杂度高；不太适合高维稀疏特征，通常采用稠密的数值特征。

参考链接：

1）https://blog.csdn.net/a819825294/article/details/51188740

CTR预估模型

softmax函数的定义是什么？

神经网络为什么会产生梯度消失现象？

常见的激活函数有哪些？都有什么特点？

挑一种激活函数推导梯度下降的过程。

Attention机制是什么？

NN

RNN

聚类算法

模型评估

GAN

LSTM

Online learning

Embedding方法

Reinforcement learning

工具

将machine learning知识应用于实际业务的工具。

Spark

Tensorflow

Xgboost

Parameter server

逻辑

举一反三的能力，解决问题的条理性，发散思维的能力。

业务

深入理解所在行业的商业模式，从业务中发现motivation并进而改进模型算法的能力。

资源

数据可视化：<http://cs109.github.io/2015/pages/videos.html>

特征工程：<http://www.cnblogs.com/jasonfreak/p/5448385.html>