作者：政化天下666  
链接：[https://www.nowcoder.com/discuss/102895](https://www.nowcoder.com/discuss/102895" \t "_blank)  
来源：牛客网

**一、机器学习**

**单模型**

1、LR的损失函数的公式和函数

2、LR的推导过程

3、LR如何解决共线性，为什么深度学习不强调

4、LR如何防止过拟合

5、LR分布式训练怎么做

6、LR为什么使用Sigmoid

7、SVM的损失函数

8、SVM的推导过程

9、SVM怎么扩展到多分类问题

10、SVM需要解决的重要数学问题是什么

11、LR和SVM的区别

12、Gini系数、信息增益、信息增益率的公式

13、CART回归和分类时节点如何划分的

14、决策树将一个特征全部乘以2会有什么影响

15、反向传播算法的推导

16、贝叶斯原理

17、L\_BFGS，DFP推导

18、Kmeans算法，如何选择k的个数

19、DBSCAN介绍

20、GMM算法

21、UBM-GMM模型

**集成学习**

1、Boosting 和Bagging的比较

2、XGB的推导

3、XGB为什么要用二阶信息不用一阶

4、XGB的VC维

5、LGB、XGB的区别和联系，并行是如何并行的

6、GBDT的原理，以及常用的调参的参数

7、XGB与GBDT的比较

8、RF怎么进行节点划分

9、GBDT和RF的比较

10、Stacking方法

**特征工程**

1、如何判断特征的有效性

2、特征选择的几种方法

3、为什么要做数据归一化，在梯度下降时有什么好处

**评价指标**

1、评价指标及含义

2、AUC理解和计算方法

3、样本分布不均衡时，怎么训练怎么评价

**损失函数、优化函数、核函数**

1、各种核函数的比较与使用场景

2、牛顿法的原理及求解sqrt <https://leetcode.com/problems/sqrtx/>

3、SGD、Momentum和Adam的区别和联系

4、GD和SGD等的区别

5、各个损失函数的形式与区别

6、交叉熵损失公式及推导

7、偏差和方差的区别

**正则化、降维、过拟合**

1、L1和L2的区别与应用场景

2、各个模型如何防止过拟合

3、使得|x\_1 - x\*| + .. + |x\_n - x\*|最小的x\*

4、SVD在遇到数据特别多的时候会产生一定的问题？如何解决？

5、PCA的原理

6、PCA与SVD之间区别和联系

**学习链接**

<http://www.dscademy.com/supervised-learning/linear-regression/>

<https://www.jianshu.com/p/70e04c02985c>

**二、深度学习**

**CNN相关**

1、各个CNN模型之间的比较，例如inception、VGG、Resnet等

2、CNN的模型结构与原理

3、Pooling的作用

4、Dropout的理解

5、BN原理及为什么可以工作

6、Resnet的原理

7、胶囊网络的原理

8、Alphago的原理

9、Data Augmetaion方法

10、1X1卷积核的作用

**RNN相关**

1、LSTM的结构、原理及参数数量

2、梯度消失原因，解决方法，为什么LSTM可以避免梯度消失

3、GRU与LSTM的不同

4、RNN模型的演变过程

5、RNN中的正则化方法：AR以及TAR

**模型比较、训练**

1、深度学习中的过拟合，如何解决

2、梯度消失梯度爆炸的原因及解决方法

3、模型训练停止方法

4、RNN和CNN的对比

**学习链接**

<https://www.cnblogs.com/huanyi0723/p/8470866.html>

**三、推荐算法**

1、FM模型的具体公式，FFM的改进

2、个性化推荐的常用模型

3、[https://www.jianshu.com/p/99e8f24ec7df](https://www.jianshu.com/p/99e8f24ec7df" \t "_blank)

**四、NLP相关**

1、Word2Vec原理，CBOW和Skip-gram的训练过程，负采样

2、TF-IDF原理

3、Fasttext的改进

4、Sentence-embedding模型

5、Doc-embedding模型

6、情感分析怎么做

7、序列标注的理解

**五、概率论**

1、如何衡量两个分布的相似度

2、CRF推导

3、统计中的P值和Alpha值

4、常问问题：摸扑克牌、硬币、五福的期望等

**六、框架**

1、Hadoop，Hive，Spark相关理论

2、Tensorflow的图计算模型

3、with关键字

4、模型保存的相关描述

5、session是什么

**七、语言**

1、C++里面的多态

2、Python numpy经典题目

3、zip作用

4、python异常处理

5、常见的知识点（实现一个0、1过滤器等简单手撕代码题）

**八、其他基础**

复习其他领域相关的知识可以参照：<https://github.com/CyC2018/Interview-Notebook>