1.kmeans和dbscan最本质的区别是啥

2.聚类中的距离度量是什么，还有其他的距离度量方式吗，可以把kmeans中常见的平方距离替换成曼哈顿距离或者余弦相似度吗？（我答肯定可以），那把平方距离换成任意其他一种距离度量方式可以吗，可以或者不可以都说说原因。如果换成余弦相似度的话怎么去用，比如说在聚成一类后怎么测算聚类中心点。把平方距离替换成其他距离公式的话会有什么问题。

3.还有哪些聚类方法？其他的聚类方法给讲一下？

4.卷积神经网络和传统全连接网络的loss有区别吗，怎么参数更新。

5.pca和自动编码器的区别是啥，是原理不一样还是效果不一样。

6.随机森林是什么树，gbdt是什么树，gbdt可以是分类树吗，gbdt的分裂标准是啥。

7.给你3个样本，只有一个特征，三个样本的特征值分别为1、2、3，对应的y值分别为4、5、6，现场建一颗gbdt的树，讲讲怎么进行分裂节点的选择

8.怎么选取最优分割点，比如有一亿个不同的样本取值点，难道你要选一亿个点作为候选分割点吗？

9.随机森林是怎么做的，包括样本选择和特征选择，详细说一遍

10.手撕随机森林的代码

11.如何设计一个深度学习网络，深度和广度怎么权衡，深的网络有什么好处

12.网络设计时全连接层一般为几层

13.如何调参，根据自己平时调参的经验来说说

14.特征工程怎么做

15.归一化和标准化有什么好处

16.场景题目，判断一个斗鱼用户的性别，你可以获取性别之外的所有信息