1，overfitting and underfitting的原因。

1）产生过拟合的原因：

训练集中的数据抽取错误，太少，或者不均衡，不足以有效代表业务逻辑或场景；

训练集中的数据噪音（异常值）干扰过大；

训练模型的“逻辑假设“到了模型应用时已经不能成立

参数太多，模型复杂度太高；

特征量太多，模型训练过度，比如决策树模型，神经网络模型

解决方法：

减少特征数量

正则化

增大样本训练规模，采样均衡

简化模型

交叉验证

去除异常值

Dropout

2）产生欠拟合的原因：

模型复杂度过低

特征量过少

解决方法：

增加新特征

增加模型复杂度

减少正则化系数

2，已经用到ML的三个场景

淘宝，siri，无人机

1. 知乎，微博，头条，豆瓣等个性化推荐
2. 公安武警刑侦办案，打击犯罪团伙
3. 证券投行金融行业量化投资

3，三个尚未应用但可以应用ML的场景

1. 反腐（判断是否腐败）
2. 城市规划（修不修地铁，地下隧道建在哪，居民区，办公区，休闲区怎么分配……）
3. 根据个人成长经历和兴趣爱好，进行职业推荐以及学习路径推荐