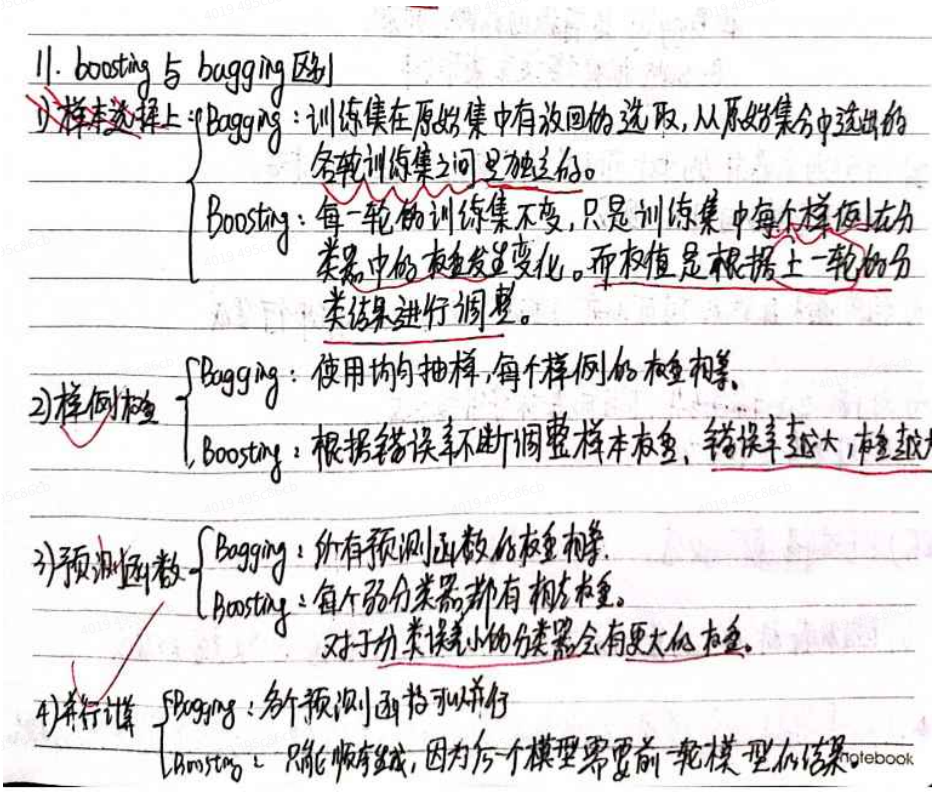
8.bagging与boost对比



B. 随机森林与GBDT区别：

1) 训练集的选取：随机森林采用Bagging思想，GBDT采用Boosting思想；

这两种方法都是Bootstrap思想的应用。

（从个数N的样集中有放回的抽取N次，每次抽1个）

但Bagging是有放回的均匀抽样

Boosting根据分错误率来取样

2) 组成随机森林的树可以是分类树，也可是回归树。

GBDT只能由回归树组成

3) 组成随机森林的树可以并行生成；而GBDT只能串行生成

4) 对于最终的输出结果，随机森林采用投票法：

GBDT采用加权累加。

5) RF对异常值不敏感，GBDT对异常值敏感

6) 随机森林对训练集一视同仁；GBDT是基于权值的弱分类器的集成

7) RF是通过减少模型方差提高性能；GBDT是通过减少模型偏差提高训练