一位老师，一位领导，一个让全体学生考上目标学校的故事

原创 夕小瑶 夕小瑶的卖萌屋 2017-04-03

今天，小夕给大家讲一个故事…

从前，有座山…

山里，有座学校…

学校里，有一位老师，一位领导，还有五只可爱的小仙(学)女(生)。

这5个学生的名字是：小兔，小青，小路，小音，小夕。她们的高考目标依次为清华、清华、清华、清华、浙大。但是不管她们的目标如何，在家长的逼迫下：

1. 假设过了浙大的线，但是没过清华的线，那就上浙大。(所以想考清华的学生会很难过)
2. 假设如果过了清华的线，那么就必须要上清华，哪怕目标是浙大，也不能报。(所以想考浙大的学生会很难过)

然后老师的培养目标呢，并不是让所有的学生都上清华，而是让每个学生都达成她们心中的目标！想上浙大的学生，就不能考的太高，以免上了清华。想上清华的学生，就要考的足够高，以免上了浙大。

1．这个老师很奇怪，他每天会制定一个精力分配计划表，比如老师每天都有100份精力，他要给各个学生分配精力。开学第一天，由于他对这些学生都不熟悉，于是他给这5个学生平均分配精力，即每人都能得到20份精力。

2．这个老师白天怎么上课呢？他会重点关注和培养今天的精力分配计划中那些精力分配很多的学生。他这一天，会讲很多次课，每节课结束的时候呢，都会安排一个随堂考试。并且根据这场考试后各个学生的成绩单，来评价一下这场考试的**成功率**。

2．但是注意啊，怎么衡量这个考试成不成功呢？重点来啦，这个老师很奇葩！他主要是看**今天重点关注的学生**能不能考上目标学校！比如今天老师的100份精力中，92份给了小夕，剩下的8份平分给了小兔等4人。那么！今天！只要小夕能达成目标(即恰好考上浙大)，那么其他4人都没有达成目标也没事(￣▽￣)，这时的考试成功率是92%。但是如果这场考试中小夕没有达成目标(即不小心考上了清华)，其他4人都达成了目标(即考上了清华)，就认为这场考试的成功率只有8%！(我仿佛听到有人说这个老师是不是智障…)

2．然后呢，按照上面讲的这么不公平的考试结果评价准则，选出今天考的最成功的一场！然后将这一场考试作为今天最终的考试结果，并且彻底忽略掉今天的其他考试。并将这场考试的**考试卷**和**成功率**连夜汇报给上级。

3．上级是一个夜猫子，而且秉公执法却不做复查的人。于是，他会在明天到来之前，根据今天老师递交的考试结果中的成功率，来给这场考试的考试卷打个**分数**。日后拿出这张考试卷时，看到打的这个分，就知道这场考试卷的含金量是多少啦。

4．老师也不敢睡觉，连夜等待上级给这场考试打的分数。等来上级打的这个分数后，老师赶紧拿着这个**分数**和**今天的精力分配计划表**来制定**明天的精力分配计划表**。

4．怎么制定呢？上级打的分数决定了精力分配计划表的变动大小，如果上级给打了0分，就意味着计划表不要变动了。上级给打的分数越高，明天的计划表就要改动越大。那么怎么改呢？这时，老师终于开始考虑全体学生了，

老师会将今天这场考试中达成目标的学生的精力分配减少（比如今天在小夕身上分配92份精力，小夕达成了目标，那么根据上级的打分，明天分配的精力要降低20%，所以明天给小夕的精力暂时记为92*(1-0.2)=73.6。），将今天这场考试中未达成目标的学生的精力分配增加（比如今天在小兔身上分配2份精力，那么根据上级的打分，明天分配的精力要增加20%，所以明天给小兔的精力暂时记为2*(1+0.2)=2.4。）。

4. 然后都算完后，发现明天给各个学生的精力之和不等于100了，那就归一化一下，保证精力之和是100。

5. 于是，明天又是崭新的一天。。。除了精力分配计划表改变了以外，其他并没有影响，完全重复前面的过程。

5. 就这样过了很多很多天。。。

6. 什么时候结束呢？

6 . 那就是根据上级对每天的每场考试的打分，来累加每个学生在前面每天的考试结果啦（即加权的考试结果，对于上级打分低的考试，考试结果就不太重要；对于上级打分高的考试，考试结果就很重要）

6 . 如果有一天，累加每个学生的考试结果后发现，诶？所有学生的累计考试结果全都达成了目标！！！即累加起来后，发现学生全都考上了目标学校！！！好了，你们毕业惹~

7. 这时，又跑过来一个没有考试目标的学生，叫小好，她问老师，“老师老师，你说我将来会上清华还是浙大呢？”

7 . 老师：“很简单，你去把这一摞卷子全都做一遍，这是我们的往日考试卷。然后你全做完后，我给你按照上级以前对每张卷子含金量的打分，给你累加一下，得到最终的考试结果。这个最终的考试结果就代表了你将来会上的学校。”

是的，上面的过程，就是机器学习的AdaBoost算法/分类器。包括了Adaboost的训练过程和分类过程。

Adaboost是集成机器学习中的典型算法，是Boosting思想的一个具体实现。通过训练并组合很多弱分类器，来加权决定分类结果。

怎么把上面的故事转成Adaboost算法呢？完成下面的概念替换：

下面映射的是算法中的实体（变量）

- 每个有目标的学生 -> 一个训练样本x_{train}
- 没有考试目标的学生 -> 待分类样本x_{test}
- 清华、浙大 -> 本任务的两个类别
- 每天的精力分配计划表 -> 每轮迭代的样本权重向量D
- 每天选出的成功率最高的考试卷 -> 每轮迭代产生的弱分类器
- 老师计算的每场考试的成功率 -> 训练样本集的加权错误率
- 上级领导给每个考试卷打的分数 -> 每个弱分类器的决策权重α
- 经过的天数 -> 弱分类器的数量

下面映射的是算法中的过程（与故事中每一段前面的序号完全对应）

- 1. 开学第一天，平均分配精力给每个学生 -> 样本权重向量D的初始化
- 2. 每一天的教学、随堂考试、选出成功率最高的考试过程 -> 本轮迭代中训练弱分类器的过程
- 3. 上级领导给考试卷打分的过程 -> 计算当前这个弱分类器的决策权重α

- 4.老师制作明天的精力分配计划表 -> 计算下一轮迭代时的样本权重向量D
- 5. 一天天的过去 -> 不断重复上述步骤2-4.
- 6. 所有学生累计考试结果全都达成目标 -> Adaboost模型收敛，完成训练。
- 7. 给没目标的小好预测学校 -> 利用训练好的Adaboost分类器预测待分类样本的类别。

听说，写Adaboost的代码的时候，在注释中把小夕讲的这个故事写出来就不会出错了...

最后，请手机/电脑/IPAD/投影仪屏幕前正在带学生的老师，请勿模仿本文
(￣∇￣)
