

QG 工作室暑期实习生两日结

姓名： 许继元 组别： 数据挖掘 年级： 19 级 日期： 2020.08.06

| | |
|----------|---|
| 生活小记 | <p>8.5 日，上午学习了 KNN 算法，它的工作机制还算比较简单，只需给定测试样本，基于某种距离度量找出训练集中与其最靠近的 k 个训练样本，然后基于这 k 个“邻居”的信息来进行预测即可。相对于前面的算法，KNN 还是挺简单的。后面还了解了一下 MDS 这种经典的降维方法。下午完成了 KNN 算法的推导以及代码实现，并在 <code>dating</code> 数据集上测试了 KNN 算法。晚上是 QG 技术交流会，是嵌入式组的师兄进行分享，师兄介绍了物联网相关的知识，了解到学习物联网方向需要学习的方向很多，基本上是步入全栈了。后面师兄还介绍了他们开发的项目，了解到了嵌入式组的强大。技术交流会过后就是康乐活动了，我和 19 级的小伙伴以及 18 级的师兄师姐一起玩了王者荣耀，这是第一次正式的线上康乐活动，不得不说还是挺好玩的。晚上睡觉前依旧抽出时间练练琴。</p> <p>8.6 日，估计是昨天康乐活动太上头了，早上晚起了一会儿，然后赶紧进会议学习，上午学习了 K-Means 算法，K-Means 算法也是一种思想简单的聚类算法，上午实现了其代码以及在 <code>iris</code> 数据集上测试了 K-Means 算法。下午学习了 NaiveBayes 算法，重新学了一下贝叶斯决策论、极大似然估计，推导了 NaiveBayes 算法，并在垃圾邮件分类数据集上测试了 NaiveBayes 算法。晚上了解了一下半朴素贝叶斯分类器、贝叶斯网以及 EM 算法。不得不说学习这些需要概率论基础的算法还是比较吃力的。睡觉前依旧抽出时间练习吉他。</p> |
| 学习开发比赛概要 | <p>学习上，学习并实现了 KNN 算法，其是“懒惰学习”的代表——在训练阶段仅把样本保存起来，训练时间开销为零，待收到测试样本后再进行处理。了解了 MDS 降维方法。学习并实现了 K-Means 算法。学习了贝叶斯决策论、极大似然估计以及 NaiveBayes 算法的实现。了解了半朴素贝叶斯分类器、贝叶斯网以及 EM 算法。</p> |

| | |
|----------------------|--|
| 感想收获 | <p>这周的前几天学的其他东西比较多，所以对于数据挖掘理论学习进度后期有点赶，这两天就赶紧加快速度完成进度了。</p> <p>技术交流会还是蛮有趣的，可以看看其他师兄师姐对于他们方向的经验分享，自己也可以从中学到一些，或者说开阔视野。</p> <p>康乐活动很好玩，和工作室小伙伴以及师兄师姐开黑很快乐。</p> |
| 存在问题 (备注) | |