

QG 工作室周记

姓名：王伟杭

组别：数据挖掘组 年级：2020 级

周次：第 4 周

生活随记	这几天生活比较日常，无聊的时候打打 LOL，跟朋友出门吃了几顿，麦当劳 yyds。
学习 开发 比赛 情况	<ol style="list-style-type: none">1. 聚类算法 K-means2. 降维算法-PCA、SVD3. 集成学习-Boosting、Bagging4. 深度学习框架入门 Pytorch、Mindspore
一周总结	<p>这一周，主要是对中期考核的存在的问题做出调整，学习一些简单的算法。</p> <p>考核结束后，先是学习了 PCA 算法和 SVD 算法，这两个算法相对来说比较简单，与线性代数练习紧密，涉及到一些矩阵分解的知识之前都有看到过，所以学起来的话比较轻松一点，也能够很好的复现算法出来，和 sklearn 的效果相比几乎一致；比较可惜的是没有把自动选取 k 值设计出来，看到了一些运用主成分来自动选取 k 值的博客，不过因为时间原因并没有动手实践。</p> <p>接下来用了几天的时间在学集成学习，AdaBoost、XgBoost，从最开始的 boosting 看起，中间老是有一些地方绕弯绕不过来，两天看了很多博客，终于实现出了一个 Adaboost，Adaboost 实现卡得最久的地方居然构建一个简单的分类器，怎么将权重考虑进去我用了很多时间去思考这个问题，好在最后在 github 上找到了答案，成功构建出了一个简单的阈值分类器，顺利构造出 Adaboost，与 sklearn 想比起来性能不相上下，就是在时间复杂度上要明显劣于 sklearn，之后就是对 Xgboost 的学习，而因为我之前对回归树还没怎么去了解，为了能够继续推进下去，我又转头去学习回归树的相关资料，花费了比较多的时间，然后才转向 Xgboost 的学习。</p> <p>对 Xgboost 的学习并不是非常顺利，看了很多的博客视频也没非常清楚的弄明白他是怎样的一个原理，为了进度考虑，遂放弃对其进行代码实现。</p> <p>之后就是简单实现一个 k-means，这个在很久以前就实现过一次，所以这次实现起来还是比较顺手的，从效果上来看与 sklearn 效果也比较接近，对球形数据聚类效果也比较可观，算是一个简单算法。</p> <p>后面的两天就是对深度学习框架的一个入门，说实话在这之前完全不懂怎么做深度</p>

	<p>学习，看见身边很多的朋友都已经在炼丹、调参，弄一些很高大上的网络，我对这样是比较抗拒的，我还是觉得应该先把基础打牢固，之前在做中期考核的时候就已经体会到了调参的快乐，完全不懂这个模型的底层都能调出很好的结果，这不是我想要的，于是这几天还是规规矩矩的学习了一下 Pytorch 和 Mindspore 这两个框架，Mindspore 的教程比较少，为了比较好的学习，我还是选择尽量以 Pytorch 为主，然后在举一反三来学习 Mindspore。但是 pytorch 的学习也不是非常的顺利，总是被教程里突然调出各种的类搞得很晕，不知道各个类的功能，我觉得这算是一个开始学习的必经之路吧，后续也学习一些比较简单的使用，知道了怎么 pytorch 是怎么使用 MNIST 来搭建模型的，也简单实践了一下做了一个线性回归、逻辑回归和手写数字识别，后面再慢慢深入吧。</p>
<p>存在问题 未来规划</p>	<p>1.偶然间看到了一些机器学习的面试题，发现自己对一些简单模型的了解学习还非常不足，以后要继续加强对算法原理的学习。</p> <p>2.加强对深度学习框架的学习，为绿色计算做准备。</p> <p>3.总结中期考核的教训，全力做好最终考核。</p>
<p>导师评价</p>	