

QG 工作室周记

姓名： 许继元

组别： 数据挖掘 年级： 19 级

周次： 第 4 周

生活随记

安装和配置了 Redis 数据库，并学了 Redis 数据库的基础内容，例如简单的命令以及数据类型等，接着安装了 Redis 数据库的可视化工具 Another Redis Desktop Manager。在 GitHub 上找到一个 star 最多的 IP 代理池，调度程序和 API 服务都编写好了，clone 下来之后配置到本地，以便后续爬虫知识的学习和实践。

在阿里云领了半年的免费服务器，计划把中期考核的项目部署上去，在服务器的 IIS 添加完网站，配置好安全组之后，出现了一个很奇怪的问题，在本机访问公网可以访问到前端，但是后台无法运行；而在服务器访问 127.0.0.1 的时候，也是无法运行后台，奇怪的是，在服务器访问 localhost 就一切正常，自己也对 0.0.0.0、127.0.0.1 以及 localhost 的概念不是很清晰，所以就先放着了。

8.4 日，完成了 BPNN 模型的数学推导以及算法代码实现，在手写字体数据集上进行了测试。8.5 日，完成了 KNN 算法的数学推导以及代码实现，在 dating 数据集上进行测试，后面还了解了一下 MDS 这种经典的降维方法。

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。

8.6 日，学习了 K-Means 算法和 NaiveBayes 算法，重新学了一下贝叶斯决策论、极大似然估计，顺便通过西瓜书了解了一下半朴素贝叶斯分类器、贝叶斯网以及 EM 算法。

8.7 日，有一个朋友也是学大数据方向的，来找我一起打阿里天池的新人实战赛，于是就开始了准备，由于可以使用框架，也想趁机了解一下 XGBoost 框架的基本使用。比赛是 o2o 优惠券使用预测，数据内容为用户在线下以及线上的领券以及消费行为记录。从数据预处理到特征工程、划分数据集、对数据集打标签。基于这些工作，接着构建 XGBoost 模型，调整好参数之后按比赛要求整理好结果文件的输出，最后在阿里云上提交了结果文件，AUC 是 0.72 左右，虽然模型有点过拟合，不过总算是完成了这次比赛。

8.8 日，学习了 HMM 这一著名的有向图模型，之前就有看过 HMM 的介绍，学起来还是蛮抽象的。下午了解了 MRF 和 CRF 这两个无向图模型，不过还没有实现其代码。晚上技术交流会结束后就是康乐活动了，还是和上次一样，和小伙伴以及师兄打王者荣耀。

学习 开发 比赛 情况	<p>学习上，简单学习了 Redis 数据库并配置了 IP 代理池。学习了部署项目到云服务器。学习了 BPNN 模型、KNN 算法、MDS 降维法、K-Means 算法、NaiveBayes 算法、半朴素贝叶斯分类器、贝叶斯网、EM 算法、XGBoost 模型、HMM 模型、MRF 模型以及 CRF 模型。备考 CET 6。</p> <p>比赛上，报名并成功提交了阿里天池新人实战赛。</p>
一周总结	<p>这周花了一部分时间进行广泛的学习，拓展自己的知识面，并参加了阿里天池新人实战赛练习数据挖掘典型题目。对于机器学习算法也有复盘和学习新知识。对于 CET 6 的备考强度还不够。</p>
存在问题 未来规划	<p>存在问题：</p> <ul style="list-style-type: none">①学习新的机器学习算法不够多。②对于 CET 6 的备考强度还不够。 <p>未来规划：</p> <ul style="list-style-type: none">①增加知识储备，准备最终考核。②备考 CET 6 增加强度。③继续学习爬虫知识。
导师评价	