



中国石油大学 (华东)
CHINA UNIVERSITY OF PETROLEUM

《计算机科学导论》课程总结报告

学生姓名：刘禹铄

学 号：2003020218

专业班级：计算 2003

学 院：计算机科学与技术学院

课程认识 30%	问题思考 30%	格式规范 20%	IT 工具 20%	Latex 附加 10%	总分	评阅教师

2022 年 11 月 15 日

1 引言

计算科学导论作为计算机科学与技术专业的第一堂课，是本专业新生未来四年专业学习的风向标，对学生如何正确认识计算科学有着指导意义。本文将从一名大三计算机专业学生的视角，结合自身经历，阐述大学生涯过半后再修读本课程对计算科学产生的新的认知与感受，以及对自身未来专业学习、发展规划的再思考。

2 对计算科学导论这门课程的认识、体会

短短八周的计算科学导论课程的学习已告一段落，未来的专业学习乃至职业发展仍在继续。在孙老师对计算科学的介绍中，我看到了计算科学的前世今生以及未来发展的无限可能。我认为这门课程最大的意义在于可以帮助初学者正确认识本专业的特性并能对未来四年的专业学习方向有一个较为清晰的认知。我个人认为，这门课会对学生产生较为深远的影响，只不过这种影响是潜移默化的，以至于有的同学会将其冠以“水课”头衔，着眼于既得利益，觉得上课时间不如完成其他课程的任务。也许这门课程讲的内容不能帮助你解决任何一道编程题，但当你未来处于某个关键节点时，你可能会想起课上老师讲的某一句话，这句话可能会影响到你的选择，乃至改变你的一生。下面我将从课程本身以及对我个人的影响两个角度出发，谈谈我对计算科学导论这门课程的认识以及体会。

首先，本课程的开设是有必要的。对于一个初入大学有理想、有抱负的计算机科学与技术专业的学生而言，树立一个正确的价值观比学习任何专业知识都重要。近几年计算机专业被神化的很严重，一度成为众多大学的热门专业，也是转专业时期的烫手山芋。在万花筒般的计算机科学与技术世界，初学者很难不受社会各种思潮的影响，可能很多同学来到这个专业不知道要做什么，说的功利点，可能为了毕业后挣钱多，以至于对本专业的学习没有一个清楚的认识，停留在对计算机专业的片面理解中。在专业学习中经常出现对课程规划的众多质疑，比如为什么计算机专业要学高等数学，为什么要学大学物理，为什么编程课那么少等，更有甚者扬言上不上大学对自己未来的工作没有任何影响，觉得大学学的东西自己以后根本用不到。我认为，如果认真修读本课程的话，这些问题都可以找到答案。一个外行毕业后学习几个月的编程语言后也能当程序员，但显然和科班出身的、拥有比较合理的专业知识结构和科学的思想方法的学生有着不小的差距。我们绝不能采取实用主义的态度学习计算机科学与技术知识，满足于能够操作使用计算机系统来处理一些各行各业的简单问题，我们不仅要知其然，更要知其所以然。

其次，本课程的教学方向是正确的。开设一门导论课，尤其是要面向整个专业的学习，要考虑很多层面，因为对于初学者而言，并没有对计算机科学与技术学科科学的认识基础，很多的专业知识讲起来会略显苍白，但如果对这些知识闭口不谈，本门课程也就失去了指导意义。我认为在这个方面做的还是很好的，本课程并没有以大幅的专业知识开篇，而是选择从科学哲学的思想方法的角度出发，奠定一个主基调，再逐渐对计算机科学与技术的定义、范畴、特点、基本问题、发展主线、学科分类、知识组织结构、学科发展的特点和内在规律等方面进行系统的、全面的阐述。课程中涉及到的很多知识可能并不重要，重要的是能学到其中的思想，以及解决问题的抽象方法。课程形式也较为灵活，这样一种较为开放性的课程并没有采取考试的形式进行考核，而是以小组汇报以及课程总结报告等多种形式作为评估标准，符合学生综合素质培养的要求。

对于个人而言，作为一名大三的学生，本课程的学习像是对自己两年专业学习成果的一次综合评价，是对过去的总结，也是对未来的展望。在此次课程的收获中，有喜，也有忧。

喜的是我在专业学习中的一些价值观得到了肯定。其中，我一直坚持的一个想法就是计算机离不开数学，或者说计算机的专业学习不能脱离理论，这可能也与自己在数学方面的兴趣有关。我经常和身边的同学争论这一话题，他们总是说自己以后搞开发根本用不到数学，也没有必要学习理论。17 级本课程的名称为《计算机科学导论》，后来改为了《计算科学导论》，这一变化就说明了理论研究的重要性。本课程的第一节内容就是计算机的数学起源，从丢番图方程到可计算性问题，揭示了计算机和数学间的种种联系。我现在仍记得孙老师上课讲的一句话“你们做编程题时，代码过了测试样例并不能说明你的代码就是百分百正确的，只能说明对这些测试样例是正确的”，当你提出一种新的算法时，如果不能从理论意义上证明这个算法的正确性，就算用一千种模拟实验验证了它的正确性，它也是不完善的。一个典型的例子就是老师上课提到过的四色定理。计算机证明虽然做了百亿次判断，终究只是在庞大的数量优势上取得成功，它的理论证明仍然是未解之谜，也成为了世界近代三大数学难题之一。

忧的是曾经自己走了很多弯路，过去的自己也做的不够好。当时填报志愿时第一志愿便选择了计算机，当时想的比较简单，可能就比较喜欢玩游戏，由于分数原因被调剂到了其他专业。上大学以后当我第一次用 C 语言解决一道编程题时，我便喜欢上了那瞬间的成就感，后来也一腔热血地来到了计算机专业。或许当时的选择并没有错，但没有做好转专业后的学习规划。当我在上课时看到老师 ppt 上展示的课程先修后继关系时，我便想到了当时修读很多课程之前并没有修读其先修课程，导致当时部分内容听起来比较吃力，影响到了专业学习。“大干四周，搞出个 CPU”，这是老师上课反复说的一句话，说的是 22 级同学大三的暑期实习。听到这句话的时候我就在想，半年多后的自己有没有这个能力，实现从逻辑 CPU 到物理 CPU 的飞跃，至少根据目前计算机组成原理课程的学习情况来说还远远不够，这也算是对自己后期学习的一种激励吧。

我个人比较喜欢进行数学方面的理论研究，也对计算科学的专业学习始终持以要知其然，更要知其所以然的态度。基于自身兴趣，我在大二学年在自己感兴趣的领域进行了一些实践学习，也算一种现身说法，进一步说明我对计算科学的认知。

2.1 大数据

大数据一词在近几年比较火，比如某些短视频平台根据个人喜好推送相关短视频，或者是大数据筛查便知道自己的旅居史。老师上课也提到过，地学院的专业数据很多，但这并不算大数据，只能说是数据大。关于大数据的定义，并没有一个官方的说法，结合我自身的经历而言，大数据一定是基于数据，但关键不是数据大，关键是处理数据的方法，以及如何让数据创造更多的应用价值。机器学习是和大数据比较相关的方法，其本质是模型的选择和参数的确定。更多的是找到一种处理目前已有数据的方法以更好地进行回归分析，并将拟合的模型应用于其他领域，进行相关行业未来预测等。

初次接触大数据方面的学习是变量选择。其最基础的模型就是线性回归模型，当时对于线性代数的基础知识掌握得还不够扎实，以至于在模型的一些理论细节方面遇到了些困难，它是怎么用向量矩阵形式简洁地表示出来的便揣摩了很久。印象比较深刻的一个问题是很多文章都说当变量的维数大于特征数时一定无法用一般线性回归模型进行拟合，但没有给出详细的解释。

这个问题也问了很多老师，最后得出了答案，定义自变量矩阵 X 是 $n \times p$ 的，其中 n 为样本数， p 为样本维数，那么协方差矩阵 $X^T X$ 是 $p \times p$ 的，而 $R(X^T X) < R(X) < n < p$ ，也就是说 $X^T X$ 是不满秩的，那么它就不可逆，导致一般线性回归模型系数向量的解 $(X^T X)^{-1} X^T y$ 无法表示，也就无法求解，对于此情况的处理方法可能就是增加一些惩罚项对变量进行降维。初学线性代数的时候根本不知道它有什么用，事实上这个问题就是线性代数的基本知识的一个应用，在当时我便明白了计算机专业的学生进行数理基础知识学习的重要性。

关于变量选择印象比较深刻的一篇文章是《Model-free variable selection in reproducing kernel Hilbert space》[2]，我认为这篇文章在机器学习领域是很考验数学功底的。它的核心思想是一个非常简单的数学思想，如果自变量 \mathbf{x} 的某一维特征 k 的变化不会对因变量 y 产生影响，那么 y 对 \mathbf{x} 的梯度 $\nabla f^*(\mathbf{x}) = (\nabla f_1^*(\mathbf{x}), \dots, \nabla f_p^*(\mathbf{x}))^T$ 的第 k 维分量应该是恒为 0 的。基于这个思想，我们便可以不依赖于函数模型，对导数进行估计，达到变量选择的效果。作者又将求解函数问题利用再生和希尔伯特空间中的表示定理转化为求解向量问题，将无穷维转化为有限维。其中《Learning with Kernels》[3] 一书中给出的表示定理的证明的核心思想为对于 $f(x, y) + g(x, y) + z^2$ 形式函数的寻优，其中 f 和 g 和 z 无关，当函数取得最小值时， z 必然取 0。看着好像很简单，但将形式复杂的目标函数化归到精简的数学模型并非易事。这都很好的解释了为什么计算机的学生需要学习高等数学，这里面很多思想方法都是值得推敲，值得拓展的，甚至有时候高等数学的知识还不够用。

大二国际周的时候有幸选到了韩竹教授的《Signal Processing And Networking For Big Data Applications》。虽然韩教授具体讲的一些算法我已经只有个大概印象了，但进一步理解了如何用数学思维去解决计算机领域的各种问题。

他课上说的一句话我到现在依然受用，“数据方面的问题没有通法，如果你碰到一种问题你就知道用哪种方法最好，你就真的精通这一领域了”，这是韩教授回答提问的时候用的一句话。我想，这也是计算科学领域研究需要注意的地方。还是以变量选择为例，变量选择的方法有很多，法无定法，要根据手中数据的特征决定应该使用哪种方法。当然很多时候我们我们不能找出最好的方法，只能确定哪些方法适用于这类问题，对于具体使用哪种算法，我们就要多做尝试，对比各种方法间的优劣，甚至要结合实际对原算法进行修改优化。

表 1: 变量选择方法比较

	好 \rightarrow 差			
	子集选择	NG	Lasso	岭回归
样本量小, 变量之间交互效应强	子集选择	NG	Lasso	岭回归
样本量适中, 变量之间交互效应适中	Lasso	NG	岭回归	子集选择
样本量大, 变量之间交互效应弱	岭回归	Lasso	NG	子集选择

大数据是我未来比较想研究的方向，它更像是计算机和数学的一种有机结合。我一直很看重理论方面的学习，知其然，更要知其所以然。当初学习数理课程时，身边的同学都不喜欢看定理的证明，说考试又不考证明，会用方法就行。我认为学习定理的证明有两个好处，首先就是证明用到的方法值得借鉴学习，更重要的就是只有看懂了定理的证明，了解了方法的原理，才能更好地运用它，才能具体情况具体分析，并推陈出新。

2.2 隐私保护

隐私保护也是个人比较感兴趣的领域，当然，它更多的是与其他方向相结合，在传统算法中融入隐私保护。当今时代，科学技术快速发展的同时隐私泄露的风险也在增大。如今，互联网和数据库技术的蓬勃发展使得数据收集不再仅仅是政府和统计部门的工作，来自各行各业的各种社交网站、购物网站和搜索引擎正在时时刻刻收集用户数据，这些用户数据会被公司作进一步的数据挖掘分析和利用。与此同时，这些数据包含了大量的个人敏感信息，对数据进行利用的同时将会不可避免的暴露个人隐私，导致隐私泄露。

这也应当是计算科学研究应注意的地方，在许多数据发布的应用程序中，如果数据发布者没有采取适当的数据保护措施，则可能造成大面积的个人隐私泄露。尤其是近几年，大数据的分析引起了人们的关注，随着数据重要性的增加，人们越来越喜欢利用数据来进行决策，由此带来的个人信息泄露的风险也在增加。

关于隐私保护，我接触的方向是差分隐私。有一种隐私攻击方法叫差分攻击，简单来说就是攻击者可以通过分析和被攻击者有关联的人的信息来达到窃取被攻击者信息的目的，即便被攻击者的信息被完全屏蔽掉了。差分隐私做的事就是向数据集中添加恰当的噪声，使得该数据集在剔除被攻击者信息后展现的统计规律和原数据集非常接近，以至于攻击者无法通过其他人的数据信息来推算出被攻击者是否在该数据集中，进而达到隐私保护的目的，这属于一种数据失真的隐私保护方案。

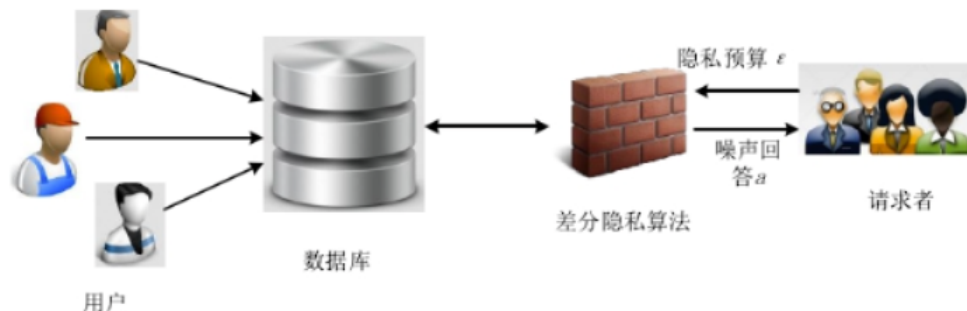


图 1: 差分隐私

在阅读相关文献的过程中，《Functional Mechanism: Regression Analysis under Differential Privacy》[4] 的作者运用数学方法解决问题的思路令我印象深刻，这篇文章首先提出了基于一般多项式形式的目标函数的差分隐私回归方法。对于向非多项式形式的目标函数的推广，他利用了我们高等数学中常见的泰勒展开，利用多项式函数近似原函数，把复杂的一般形式函数问题化归到我们较为熟悉的多项式形式函数的问题。这些基础工具看起来很简单，但当我们真正需要用其解决实际问题时，可能很难跳出固有思维，压根不会想到如此简单的数学技巧能派上用场。

差分隐私和其他算法的结合能力是很强的，这也体现了计算科学的学科交叉，很多领域、很多算法之间的核心思想都是共通的，是可以互相借鉴，彼此结合的。《回归算法中的差分隐私保护方法研究》[5] 中将传统的线性回归中的解向量作为差分隐私保护的数据，并将遗传算法融入其中，在遗传算法数据准备的过程中自然地添加噪声，利用差分隐私的指数机制对最优解向量进行选择的同时达到保护隐私的目的。

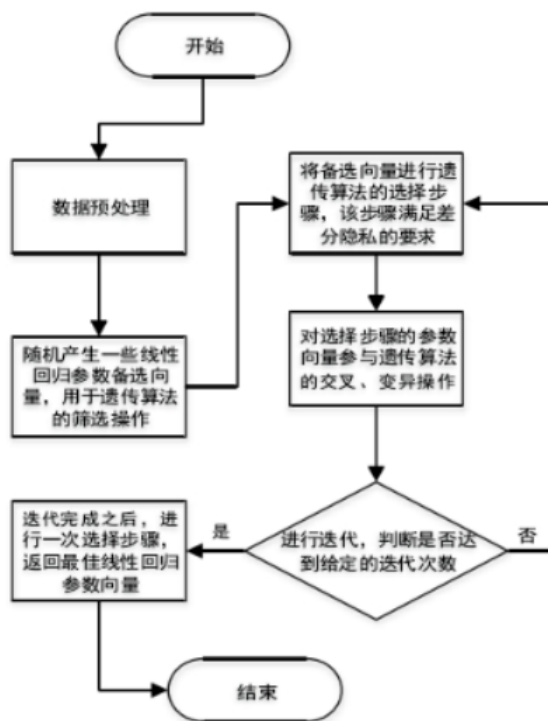


图 2: 基于遗传算法的差分隐私回归

我对隐私保护领域未来的发展抱有很高的期待。随着互联网技术、网络通信技术、数据采集技术和移动智能终端的发展促进了各种应用领域中广泛的数据收集和传输，从而存储的数据呈现了爆炸式的增长。通常在这些数据中显式的或者隐晦的隐藏着关于个人的敏感信息，而这些敏感信息容易遭受到不法分子的恶意攻击和隐私泄露的风险。也就导致了未来我们对隐私保护技术的需求会大幅增加。

3 进一步的思考

我们选择的分组演讲的题目是量子计算，这个选题是我建议的，可能老师在讲解分组演讲的要求时我们就已经确定了。我曾经在项目组中汇报过关于量子算法的内容，也读过一些相关文献，我看到了量子算法的无限可能。同时我也希望大家能正确认识量子计算，不能让其变成一门神学，这一切都要基于自己对于其中原理的理解，懂了量子计算的原理，就不会被关于量子的舆论所影响，才会有自己思考，才有可能推动量子算法的发展。

量子计算的矛盾在于其计算过程是并行的，但测量结果不是，以至于即便瞬间把结果计算出来了，有可能结果的测量会带来高额的时间成本。实现量子算法的关键就在于通过尽可能少的测量次数获得尽可能多的所需信息，也就是说要通过量子操作，改变态矢量的组合以及基矢前的系数，使我们想要的信息尽可能集中于一个态，即这个态前的系数尽可能大，甚至我们对结果测量一次就能得出结论。量子算法带来的加速效果非常显著，甚至能达到指数级的加速，如果能应用到某些关键算法中其带来的收益难以想象。Shor 算法便是求解一个可分解的正奇数 N 的素数质因子的量子算法，其相比经典算法有着指数级的加速效果。Shor 算法的出现直接威胁

到经典通讯中基于大合数分解十分困难 RSA 加密算法。

当然，不要神化量子算法，量子算法是有很大的局限性的。首先目前的量子计算机还不够成熟，只能进行中小规模的运算，而且量子态是不稳定的，会影响到算法的精度。再加上前面提到的量子算法设计的困难性，不是所有的算法都能转化为对应的量子算法，或者说目前只有很少的算法能做到，而且有的加速效果并不够明显，无法满足降低时间成本的需求。所以目前能用量子算法解决的问题少之又少，并不是所谓的“遇事不决，量子力学”。

《Quantum differentially private sparse regression learning》[1] 一文将量子算法应用于传统的 Lasso 回归，我个人认为其突出贡献在于基于 Grover 算法提出了搜索最小值的量子算法，很大程度上降低了使用 Frank-Wolfe 算法求解 Lasso 模型的时间复杂度。其另一贡献在于将量子算法和我前面提到的差分隐私巧妙地结合在了一起，就像论文中提到的一样，量子的不稳定性导致量子算法产生的噪声对绝大多数算法都会产生负面影响，因为降低了其准确性，但却有益于差分隐私算法，因为差分隐私恰好需要噪声进行隐私保护，作者找到了一种方式通过量子信道把二者联系了起来。可见量子领域还有很多值得探索、挖掘的地方，但任重而道远，每走一步都异常艰辛。

下面是我们关于小组汇报问题的进一步思考：

- 今年量子纠缠为什么能获得诺贝尔奖？

2022 年诺贝尔物理学奖授予了约翰·克劳瑟，阿兰·阿斯佩克和安东·泽林格。他们通过纠缠光子实验，确定贝尔不等式在量子世界中不成立，并开创了量子信息这一学科。

在理论物理学中，贝尔不等式是一个有关是否存在完备局域隐变量理论的不等式。实验表明贝尔不等式不成立，说明不存在关于局域隐变量的物理理论可以复制量子力学的每一个预测。约翰·克劳泽在 1972 年进行了这项实验，结果表明量子力学确实有效，而且不存在隐藏变量。然后在 1982 年，艾伦方面进行了一个更为严格的贝尔测试，它弥补了可能存在未知隐变量的漏洞。

贝尔理论 1964 年就已提出，本次实验的证明并不算是新的理论，我认为他们获得诺贝尔奖的主要原因在于对量子通信的推动。在 20 世纪 90 年代，安东·泽林格对纠缠态进行了测量，以证明所谓的量子隐形传态。即如果你取两个纠缠电子，一个与第三个电子相互作用，那么这个纠缠态就可以被转移。也就是说我们可以利用纠缠态传递量子信息，这个过程传递的是两个粒子的状态，而非量子本身。这个概念对于今天进行的许多量子通信研究来说，都是至关重要的。

诺贝尔物理学委员会主席安德斯·伊尔贝克在颁奖时说，三位获奖者各自使用“两个粒子即使在分离时也表现得像一个单元”的纠缠量子态，进行了开创性实验，实验结果为基于量子信息的新技术扫清了障碍。“对纠缠态的研究非常重要，甚至超越了解释量子力学的基本问题”。所有这三位科学家都进行了开创性的研究，为第二次量子革命铺平了道路，所以诺贝尔奖实至名归。

- 量子计算机能否取代传统计算机？

对于这个问题，学术界的主流观点是，在可见的未来，不会

首先，量子计算机只有在处理能设计出高效量子算法的特定问题时，才能超过经典计算机。对于没有量子算法或者无法找到合适的量子算法的问题，例如最简单的加减乘除，量子计

算机就没有任何优势。也就是说量子计算并不是在任何方面都优于经典计算机，只是在处理特定问题上会有奇效。所以量子计算机的出现并不会意味着经典计算机的消亡。

其次，要想让量子计算机具备我们日常生活中计算机的全部功能，比如游戏、影视等，还需要很长一段时间。通用量子计算机是一个超出目前科技水平太多的技术。以至于大多数科学家更愿意研究具有特定量子结构的量子计算机，用来执行特定的量子计算功能。比如说 Google 有一项量子计算需求，就为此配一台能专门完成这项量子计算的量子计算机就能运行的很好，搞不定的部分再交给电子计算机处理分工处理就行。即当遇到传统计算机无法解决的问题时，再考虑量子计算机。

- 从量子计算到量子计算机还有多久？

上面的回答可能已经给出了这个问题的部分答案。这取决于对量子计算机的定义是什么，如果只需要这台量子计算机执行某项特殊功能，不用考虑别的功能的话，现在就可以做到。但如果我们想让这台量子计算机服务于我们的日常生活，面临的困难前面已经提到过了，只能说道阻且长，行则将至，如果真的到了那一天，“量子计算机能否取代传统计算机”这一问题就要被重新定义了。

- 如果想学习量子计算，应该从什么时候开始？

前段时间我和北京邮电大学的同学交流时，也聊到了量子计算这一话题，从他口中我得知北京邮电大学已经开设了量子计算导论作为两学分的专业限选课。近几年中国科学技术大学也开设了量子信息专业，清华大学也成立了量子信息班，越来越多高校在注重这方面人才的培养，或引导本科生接触量子计算相关领域。

对我个人而言，量子计算是一个学科交叉性非常强的领域，比如物理、数学、计算机，而且要求你对每一个领域都非常了解。所以我认为如果想专门研究量子计算的话，计算机专业本科阶段可能不太合适，因为本科还处在一个学习基础知识，探求兴趣的阶段，这时候专门研究量子计算的话，可能在学习本专业知识的同时也要涉猎外专业很多内容，以至于专业内外的基础都可能打不牢。本科生可以先提前接触了解，等到研究生阶段，我们有更多的学科交叉的机会，而且有了较好的专业基础和综合素质时，可以考虑从事量子计算方面的专业研究。

4 总结

计算科学对我而言意味着什么，或许就是热爱。可能就是源于当初那十几行代码，源于学到的每一个算法。也许课业繁忙，也许压力沉重，但总是乐此不疲。每当老师上课提到计算科学的某一领域时，我都在问自己，作为大三的学生对这方面了解多少，能不能比之前做得更好。分组演讲的 1734 个选题拓宽了我的视野，让我看到了计算科学的更多可能，也为我提供了更多的专业兴趣选择。

悟已往之不谏，知来者之可追，我会坚定自己的信念，坚持理论研究和工程实践相结合，不负自己，不忘感恩，向两年前的自己交上一份完美答卷。

5 附录

- 注册 Github

个人网址: <https://github.com/LiuYasuo>

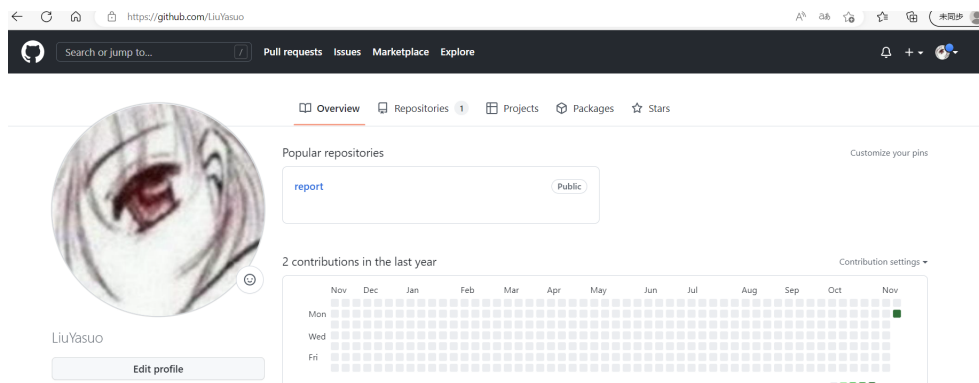


图 3: Github 个人主页

- 注册观察者



图 4: 观察者主页

- 注册学习强国



图 5: 学习强国主页

- 注册哔哩哔哩



图 6: 哔哩哔哩主页

- 注册 CSDN

个人网址: https://blog.csdn.net/qq_58678596

个人网站:



图 7: CSDN 个人主页

- 注册博客园

个人网址: <https://home.cnblogs.com/u/3036569>

个人网站:

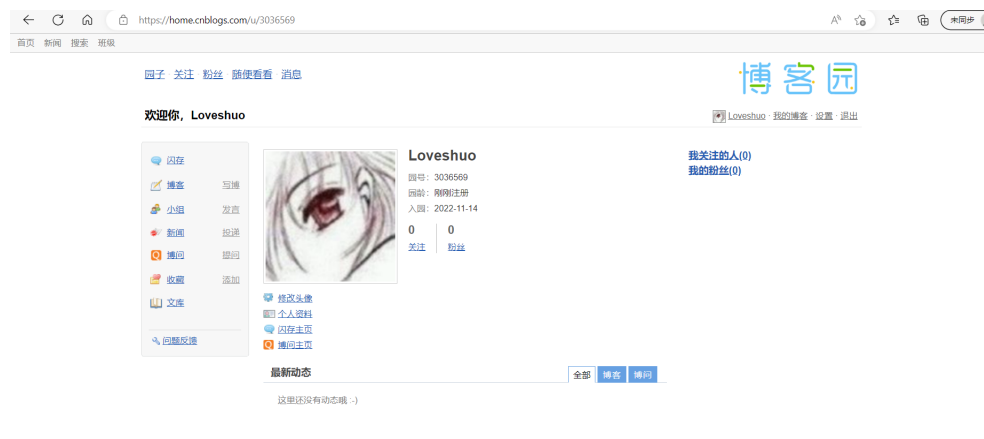


图 8: 博客园个人主页

- 注册小木虫

个人网址: <http://muchong.com/bbs/space.php?uid=32299302>

个人网站:



图 9: 小木虫个人主页

参考文献

- [1] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Shan You, and Dacheng Tao. Quantum Differentially Private Sparse Regression Learning. *arXiv e-prints*, page arXiv:2007.11921, July 2020.
- [2] Yang Lei, Lv Shaogao, and Wang Junhui. Model-free variable selection in reproducing kernel hilbert space. *Journal of Machine Learning Research*, 17(1):24, 2016.
- [3] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels*. 2018.
- [4] Jun Zhang, Zhenjie Zhang, Xiaokui Xiao, Yin Yang, and Marianne Winslett. Functional Mechanism: Regression Analysis under Differential Privacy. *arXiv e-prints*, page arXiv:1208.0219, August 2012.
- [5] 陈子彬. 回归算法中的差分隐私保护方法研究. Master's thesis, 广东工业大学, 2019.