

# 数据科学与工程专业 博士研究生培养方案

2024.9

大数据科学研究中心

国际机器学习研究中心





## 一、培养目标

面向国家社会经济建设，瞄准国际学科前沿，发挥北京大学文、理、医、工、社等学科综合优势，实现多学科交叉和产学研结合，培养学生具备坚实宽广的数据科学理论基础和系统深入的专门知识，熟练地掌握一门外国语；要求学生对于数据科学及相关学科的理论、方法与技术有透彻了解和把握，有从事本学科及跨学科研究的领悟力，理解科学研究的真谛，掌握大数据科学研究的方法；善于发现学科的前沿性问题，并对之进行深入的研究。具有严谨求实的科学态度和作风，能独立从事数据科学方面的基础研究、应用基础研究和关键技术创新等高水平研究工作；可在高等院校和研究单位从事教学和研究工作，也可在工业技术、健康医疗、环境、能源、时空、交通、生命科学、社会调查、金融与电子商务等相关领域从事专业性研究、开发和管理工作的，学生毕业后能成长为数据科学家、人工智能专家、大数据技术专家、领域大数据分析师等研究型高端人才。

## 二、研究方向

1. 机器学习的算法、理论及应用；
2. 科学智能的算法、理论及应用。

### 三、学习年限

直博生：5 年

硕博连读生：2+3 或 3+3

普博生：4 年

### 四、学位授予

一级学科：数据科学与工程 990100

专业代码	专业名称	所属学科门类	授予学位
990101	数据科学与工程（数据分析）	理学	理学博士学位
990102	数据科学与工程（应用数据科学）	理学	理学博士学位
990103	数据科学与工程（数据工程）	工学	工学博士学位
990104	数据科学与工程（数据伦理与价值）	工学	工学博士学位

### 五、学分要求

#### （一）直博生

总学分	必修	公共必修	第一外国语 2
-----	----	------	---------

40	25	5	政治理论课 2	
			论文写作课 1	
		专业必修 20	第一部分： 核心必修 9	深度学习：算法与应用 3
				科研实践（上） 2
				科研实践（下） 2
			第二部分： 分模块必修 11	教学实习 2
				统计与机器学习：至少 1 门
				计算方法：至少 1 门
				信息科学：至少 1 门
	选修 15	三个模块中任意选择		

## （二）硕博连读生

总学分 40	必修 26	公共必修 6	第一外国语 2	
			政治理论课 3	
			论文写作课 1	
		专业必修 20	第一部分： 核心必修 5	深度学习：算法与应用 3
				教学实习 2
			第二部分： 分模块必修 15	统计与机器学习：至少 1 门
				计算方法：至少 1 门
				信息科学：至少 1 门
				三个模块中任意选择
	选修	三个模块中任意选择		

	14	
--	----	--

说明：完成硕士研究生培养方案中规定的 30 学分。如硕士阶段已完成硕士生第一外国语、政治理论课和论文写作课的学习，且获得学分，则无须修读博士生第一外国语、政治理论课和论文写作课。

### (三) 普博生

总学分 18	必修 15	公共必修 5	第一外国语 2	
			政治理论课 2	
			论文写作课 1	
		专业必修 10	第一部分：核 心必修 6	科研实践（上） 2
				科研实践（下） 2
	教学实习 2			
	第二部分：分 模块必修 4	三个模块中任意选至少 2 门		
选修 3	三个模块中任意选择			

## 六、指导方式

在博士研究生培养过程中，应合理安排课程学习、社会实践、科学研究、学术交流等各方面，着重培养博士研究生的优良学风、学术探索精神、创新意识、独立科学研究能力。

导师为学生培养的第一责任人。在学生培养阶段，导师有责任对学生的选课、科研实习和论文进行指导。

中心实行轮转制。每位直博生入学后，在第一学年内根

据自己的兴趣选择不同实验室完成 2 次轮转,最后于学年末 (次年 6 月 30 日前),交回《导师接收函》,完成定导 (详见《数据科学与工程专业博士研究生轮转制实施细则》)。

## 七、课程设置

### (一) 公共必修课

序号	课程类别	课程编号	课程名称	学分
1	第一外国语	-	研究生院开设,见个人培养计划	2
2	政治理论课	61400001	中国马克思主义与当代	2
3	论文写作课	08402101	科研规范、数据分析和论文写作	1

### (二) 专业课

第一部分：核心必修				
序号	课程属性	课程编号	课程名称	学分
1	必修	08408005	深度学习：算法与应用（穆亚东，春）	3
2		08410000	教学实习	2
3		待开课	科研实践（上）	2
4		待开课	科研实践（下）	2
第二部分：分模块必修&选修				
模块一：统计与机器学习				



序号	课程属性	课程编号	课程名称	学分
1	必修	08408011	高等机器学习（王立威，春）	3
2		00104042	深度学习理论选讲	3
3		00137913	机器学习数学导引	3
4		00112640	高等统计学	3
5		00102892	统计学习	3
6		00103335	深度学习与强化学习	3
7		00102068	高维概率论	3
8		04802015	深度学习技术与应用	3
9	选修 (推荐)	00100877	贝叶斯理论与算法	3
10		00102441	统计和生物统计中的因果推断	3
11		00102442	高等深度学习	3
12		00101756	现代统计模型	3
13		00103079	理论机器学习	3
14		00113540	测度论	3
15		00102728	大数据案例实务	3
16		00112630	高等概率论	3
17		00135460	数理统计	3

模块二：计算方法				
1	必修	08403979	深度学习中的最优化方法（袁坤，秋）	3
2		08404127	单细胞组学数据分析：计算系统生物学视角（周沛劫，春）	3
3		08404334	数据科学导引 B（张文涛，秋）	3
4		00114250	机器学习（牟克典，秋，两年一开）	3
5		00100863	大数据分析中的算法（春）	3
6		00100887	最优化方法（秋）	3
7		00110860	并行计算 II	3
8		00102906	凸优化	3
9	选修 （推荐）	00112700	偏微分方程数值解	3
模块三：信息科学				
1	必修	08403574	深度强化学习方法（张行功，秋）	3
2		04812181	海量图数据的管理和挖掘（昌平校区）	3
3		04814310	数据库与知识库系统原理（李红燕，智能学院，双校区）	3
4		04804013	数据可视化（袁晓如，智能学院，校本部）	3
5		04814130	模式识别（封举富，智能学院，昌平校	3

			区)	
6		04812120	数据库原理与技术(计算机学院,昌平校区)	3
7		04802012	网络大数据管理理论和应用(计算机学院,昌平校区)	3
8		04812160	计算语言学(计算机学院,昌平校区)	3
9		04812190	网络信息体系结构(计算机学院,昌平校区)	3
10		04812102	视频编码与理解(计算机学院,双校区)	3
11		04812030	算法分析和复杂性理论(计算机学院,昌平校区)	3
12		04802052	深度生成模型(计算机学院,昌平校区)	
13		04804008	神经网络与深度学习(邓志鸿,智能学院,昌平校区)	3
14	选修 (推荐)	04804018	Python 大数据分析原理与应用	2
15		04814150	计算机视觉	3
16		04804009	图像理解	3
17		04812040	分布式系统概念与设计	3
18		04802019	计算视觉理论、模型与方法	3
19		04804014	机器学习中的优化方法	3

20		04802046	区块链技术与应用	2
21		04811982	自然语言处理高级专题	3
22		04804020	机器学习前沿	3
23		04802017	社会与市场中的计算问题选讲	3
24		04802020	机器学习原理	3
25		04833510	视觉计算与深度学习	2

说明：

1. 每学期的课程选择需征得导师的同意。
2. 研究生外语课和政治理论课，原则上应于第一学年修完。若第一外国语非英语，则必须另外选修英语作为第二外国语。
3. 专业必修课相关要求
  - (1) 第一部分：核心必修，为本专业所有研究方向的必修课，建议尽早完成。其中，“科研实践（上）”和“科研实践（下）”，须分别于第一学年上、下两学期修完。
  - (2) 第二部分：分模块必修，应从三个模块中至少选4门（每个模块至少选1门，剩下1门可在三个模块中任选）。
4. 专业选修课相关要求
 

必修课学分超出部分可代替选修课学分，因此鼓励多选必修课。

## 5. 先修课相关要求

对于本科没有修过高等数学、线性代数、统计学、数据结构与算法、数据库管理系统、等部分课程的研究生，需要在导师小组指导下先修（补修，不计学分）以上课程。选修其他院系和学校的课程，服从具体相关条例管理。

6. 成绩评定与管理参照《北京大学研究生手册》。

7. 博士研究生应在进行答辩前一学期完成全部修课要求，如未能按时完成，应及时办理延期手续。

## 八、培养环节

### （一）博士研究生资格考试(学科综合考试)

资格考试是综合考察学生的学习和科研能力、深入严肃评估学生的科研计划、把控博士研究生培养质量的重要环节。

#### 1. 完成时间

原则上所有必修课程修完后（教学实习和文献导读除外），才可以进行综合考试。

直博生：建议于第二学年第一学期完成，最晚于第三学年第一学期完成；

硕博连读生：和硕转博考试合并，不再单独举行；

普博生：建议于第一学年第二学期完成，最晚于第二学年第一学期完成。

## 2. 考核方式

由导师决定并组织学科综合考试委员会，学生进行口头汇报，答辩时长为 40-60 分钟，包括问答。按照导师所在各学院的资格考试要求，对博士研究生的学科知识、研究能力和外语水平等进行综合考察。委员会由 5 位（含）副高及以上职称的本专业和相关专业的专家组成，另设秘书 1 位，负责答辩记录和考评意见记录。导师一般不担任组长；专家所在单位应具体到二级单位，如“北京大学前沿交叉学科研究院”。

## 3. 考核结果

按合格和不合格两级评定成绩；每位学生仅有两次机会。第一次考核未通过的学生，经考试委员会同意可申请三个月后补考一次或依据考试方案参加下一次考试。对补考仍不合格者，一般予以退学；直博生和硕博连读生，由考试委员会同意后可转为硕士生培养，并报批研究生院，按学籍管理相关规定办理。

## 4. 提交材料

(1) 1 份《北京大学攻读博士学位研究生学科综合考试审

核表》（门户下载）

（2）1份综合考试答辩记录（秘书提供）

（3）导师要求提交的材料（如综考报告、读书报告等，自愿提交）

综合考试完成后，学生将上述材料纸质版交至中心办公室静园六院 203。

## （二）开题

### 1. 选题要求

研究生在广泛调查研究、阅读文献资料、搞清楚主攻方向上的前沿成果和发展动态的基础上，自己提出学位论文选题。选题应尽可能对学术发展、经济建设和社会进步有重要意义。应在规定的时间内，就选题意义、前人相关成果、材料基础与实验条件、理论与方法等方面做学位论文选题报告，尽可能广泛地听取专家意见。导师和指导小组应严格把关。博士研究生需在规规定学期初主动与导师联系开题事宜。

### 2. 完成时间

通过博士生学科综合考试后半年内。

直博生：最晚于第三学年第二学期完成；

硕博连读生：最晚于博士阶段第二学年第一学期完成；

普博生：最晚于第二学年第二学期完成。

### 3. 形式

由导师组成选题报告委员会，学生进行口头汇报，答辩时长为 40-60 分钟，包括问答。委员会由 5 位（含）副高及以上职称的本专业和相关专业的专家组成，另设秘书 1 位，负责答辩记录和考评意见记录。导师一般不担任组长；专家所在单位应具体到二级单位，如“北京大学前沿交叉学科研究院”。

### 4. 提交材料

- (1) 1 份《北京大学攻读博士学位研究生选题报告审核表》  
(门户下载)
- (2) 1 份选题报告答辩记录（秘书提供）
- (3) 1 份博士学位论文选题报告（自愿提交）

开题完成后，学生将上述材料纸质版交至中心办公室静园六院 203。

5. 选题指导小组应严格把关，帮助学生分析问题的可行性，论文工作中的难点，找出不足，明确下一步研究的方向和重点，保证论文研究工作的顺利进行。

### (三) 学位论文全面审查（预答辩）



## 1. 考核目的

听取学生论文进展情况，找出其中不足，提出后期改进意见和建议；明确下一步研究方向和重点。

## 2. 完成时间

通过开题之后，最后一学年第一学期。

## 3. 形式

由导师组成预答辩委员会，学生进行口头汇报，答辩时长为 40-60 分钟，包括问答。委员会由 5 位（含）副高及以上职称的本专业和相关专业的专家组成，另设秘书 1 位，负责答辩记录和考评意见记录。导师一般不担任组长；专家所在单位应具体到二级单位，如“北京大学前沿交叉学科研究院”。

## 4. 提交材料

- (1) 1 份《北京大学博士学位学位论文审查表》（门户下载）
- (2) 1 份预答辩记录（秘书提供）
- (3) 1 份博士学位论文预答辩报告（自愿提交）

预答辩完成后，学生将上述材料纸质版交至中心办公室静园六院 203。

## 5. 考核委员会针对学生的课题进展情况对学生进行考核，

提出评估和建议。

#### （四）年度审核

##### 1. 考核对象

中心所有在读博士研究生（包括直博生、硕博连读生和普博生），每学年均需参加一次年度考核。

##### 2. 考核时间

每学年考核时间以当年通知为准。

##### 3. 考核委员会

由 3 位及以上博士生指导小组成员担任考核委员会，导师不可担任组长。

##### 4. 考核形式

现场综合面试，每位博士生进行不少于 20 分钟的口头或 PPT 汇报。非特殊情况不得采取远程通讯的形式考核。

可参考但不局限于以下形式：

（1）第一学年年度考核可以学习情况与研究课题综述的考核方式进行；

（2）第二学年年度考核可以综合考试面试方式进行；

（3）第三年年年度考核可以开题报告方式进行；

（4）第四学年年度考核可以预答辩报告方进行；

## 5. 考核内容

针对博士生的培养环节完成情况（包括课程学习、学科综合考试、开题、预答辩等环节），以及研究进展、导学关系、心理健康等各方面进行综合评估。

## 6. 考核原则

应严把培养质量控制的关键环节，包括课程学习、学科综合考试、开题、学术成果检查、预答辩等，充分发挥学术方向引领、学习方法指导、研究内容把关、科研成果评价等方面的作用。导师及博士生指导小组应按学校要求定期对其指导的博士研究生的培养状况进行评估和情况反馈。

## 7. 提交材料

《大数据科学研究中心、国际机器学习研究中心在读博士研究生年度审核情况表》

年度审核完成后，学生将上述材料纸质版，于指定日期前，交至中心办公室静园六院 203。培养委员会将进行审核，对于没有按时完成学分或科研进度落后的同学，将约谈导师和学生，督促尽快完成培养方案的各个环节。

## 九、学位论文基本要求

博士研究生毕业学位论文必须是一篇系统完整的学术

论文，应在导师指导下，由博士研究生本人独立完成，并严格按照最新版《北京大学研究生手册》中“北京大学研究生学位论文及论文摘要的基本要求与书写格式”的规定撰写。

博士学位获得者应掌握本学科的基础理论和专业知识，熟练掌握专业外语，对本学科的研究动态及最新进展具有全面了解；具有独立从事科学研究的能力；并对所研究的问题有新的见解，取得有创造性的新成果，完成具有创新性、系统性和独立性的博士学位论文。鼓励博士研究生将学术成果以论文方式展现在数据科学领域内知名期刊、国际会议或核心学术刊物上，或以专利等形式展示。

**学术论文署名要求：**学术论文作者单位必须为北京大学/北京大学大数据科学研究中心/北京大学国际机器学习研究中心。