课程详述

COURSE SPECIFICATION

　　以下课程信息可能根据实际授课需要或在课程检讨之后产生变动。如对课程有任何疑问，请联系授课教师。

　　The course information as follows may be subject to change, either during the session because of unforeseen circumstances, or following review of the course at the end of the session. Queries about the course should be directed to the course instructor.

1.	课程名称 Course Title	地球科学大数据与人工智能 Big Data and Artificial Intelligence in Geosciences
2.	授课院系 **Originating Department**	地球与空间科学系　Department of Earth and Space Sciences
3.	课程编号 **Course Code**	ESS412
4.	课程学分 Credit Value	3
5.	课程类别 **Course Type**	专业选修课 Major Elective Courses
6.	授课学期 **Semester**	秋季 Fall
7.	授课语言 **Teaching Language**	中英双语 English & Chinese
8.	授课教师、所属学系、联系方式（如属团队授课，请列明其他授课教师） **Instructor(s), Affiliation& Contact** （**For team teaching, please list all instructors**）	陈克杰，地球与空间科学系 邮箱：chenkj@sustech.edu.cn 电话：0755-88018645 办公室：创园 9 栋 310 Kejie Chen, Department of Earth and Space Sciences E-mail: chenkj@sustech.edu.cn Tel: 0755-88018645 Office: Innovation Park #9-310
9.	实验员/助教、所属学系、联系方式 **Tutor/TA(s), Contact**	待公布 To be announced
10.	选课人数限额(可不填) **Maximum Enrolment** （**Optional**）	

11.	授课方式 **Delivery Method**	讲授 **Lectures**	习题/辅导/讨论 **Tutorials**	实验/实习 **Lab/Practical**	其它(请具体注明) **Other**（**Please specify**）	总学时 **Total**
	学时数 **Credit Hours**	36		12		48

12.	先修课程、其它学习要求 **Pre-requisites or Other Academic Requirements**	MA107A 线性代数 A、CS102B 计算机程序设计基础 B MA107A Linear Algebra A and CS102B Introduction to Computer Programming B
13.	后续课程、其它学习规划 **Courses for which this course is a pre-requisite**	
14.	其它要求修读本课程的学系 **Cross-listing Dept.**	

<center>教学大纲及教学日历 SYLLABUS</center>

15.　　教学目标 Course Objectives

　　　　本课程系统介绍地球科学大数据与人工智能的基本框架与原理，重点分析高维数据降维度、分类与预测、大图形社区结构识别、无限流数据处理、机器学习及人工智能地学的建模过程。通过学习本课程，使学生具备运用地学大数据与机器学习算法解决地球科学问题的基本能力。

This course systematically introduces the basic framework and principles of big data and artificial intelligence in earth sciences, focusing on the analysis of high-dimensional data dimension descending, classification and prediction, large graphic community structure identification, infinite stream data processing, machine learning and artificial intelligence geoscience modeling. By studying this course, students will be able to use geoscience big data and machine learning algorithms to solve earth science problems.

16.　　预达学习成果 Learning Outcomes

学生完成本课程后，将会掌握以下知识：
1. 数据清洗与预处理；
2. 高维数据降维；
3. 分类与预测；
4. 图形数据处理；
5. 无限流数据与时间序列；
6. 机器学习与深度学习；
7. 贝叶斯原理与人工智能地震学。

Upon completing the course, students will master the following knowledge:
1. Data cleaning and pre-processing;
2. High-dimensional data dimensionality reduction;
3. Classification and prediction;
4. Graphical data processing;
5. Infinite stream data and time Series;
6. Machine learning and deep learning;
7. Bayesian principle and artificial intelligence seismology.

17.　　课程内容及教学日历 （如授课语言以英文为主，则课程内容介绍可以用英文；如团队教学或模块教学，教学日历须注明主讲人）
　　Course Contents (in Parts/Chapters/Sections/Weeks. Please notify name of instructor for course section(s), if this is a team teaching or module course.)

第一章：绪论 （2 学时）
讲授 科学第四范式、地球科学大数据机器挖掘、建模

第二章：数据清洗与预处理（6 学时）
讲授 数据清洗的概念、数据集成与融合、数据变换、数据规约、离群点检测

第三章：高维数据降纬（6 学时）
讲授 相关分析、典型相关分析、哈希算法、主成分分析、因子分析

第四章：分类与预测（6 学时）
讲授 回归分析、聚类分析、判别分析、关联规则算法、推荐系统算法

第五章：无限流数据与时间序列（6 学时）
讲授 无限流数据与时序模式、数据特征提取、时间序列算法

第六章：机器学习与深度学习（10 学时）
讲授 机器学习的发展史、机器学习分类、人工神经网络、深度学习

第七章：应用实践：利用机器学习区分地震信号与背景噪声（12 学时）
实验 利用 Python 库 TensorFlow，Keras 等进行地震观测值样本训练，评价不同分类方法表现

Chapter 1: Introduction (2 hours)
The fourth normal form of science, big data machine mining and modeling of earth sciences

Chapter 2: Data Cleaning and Preprocessing (6 hours)
The concepts of data cleaning, data integration and fusion, data transformation, data protocol, outlier detection

Chapter 3: Dimension Reduction in High-Dimensional Data (6 hours)
Correlation analysis, canonical correlation analysis, hash algorithm, principal component analysis, factor analysis

Chapter 4: Classification and Forecasting (6 hours)
Regression analysis, cluster analysis, discriminant analysis, association rule algorithm, recommendation system algorithm

Chapter 5: Infinite Stream Data and Time Series (6 hours)
Infinite stream data and time series patterns, data feature extraction, time series algorithms

Chapter 6: Machine Learning and Deep Learning (10 hours)
The history of machine learning, classification of machine science, artificial neural networks, and deep learning

Chapter 7: Practise: Seismic signal/noise discrimination with machine learning (12 hours)
Using Python modules such as TensorFlow, Keras to apply machine learning to train dasets, assessing the performace of different classifiers

18. **教材及其它参考资料 Textbook and Supplementary Readings**

教材：
周永章，张良均，张奥多，王俊。地球科学大数据挖掘与机器学习，中山大学出版社，2018

参考资料：
1. Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning, Springer, 2017
2. Ian Goodfellow，Yoshua Bengio，Aaron Courville. Deep Learning, MIT Press, 2016
3. 周志华。机器学习，清华大学出版社，2016

课程评估 ASSESSMENT

19.	评估形式 Type of Assessment	评估时间 Time	占考试总成绩百分比 % of final score	违纪处罚 Penalty	备注 Notes
	出勤 Attendance		10		
	课堂表现 Class Performance				
	小测验 Quiz				
	课程项目 Projects				
	平时作业 Assignments		20		
	期中考试 Mid-Term Test		30		
	期末考试 Final Exam		40		
	期末报告 Final Presentation				
	其它（可根据需要改写以上评估方式） Others (The above may be modified as necessary)				

20.　　　记分方式 GRADING SYSTEM

> √A. 十三级等级制 Letter Grading
> □ B. 二级记分制（通过/不通过） Pass/Fail Grading

课程审批 REVIEW AND APPROVAL

21.　　　本课程设置已经过以下责任人/委员会审议通过
　　　　　This Course has been approved by the following person or committee of authority

> 地球与空间科学系本科教学指导委员会