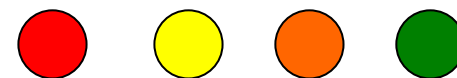


# Python程序设计与数据科学导论

## —— 课程内容回顾与总结

信息科学技术学院

胡俊峰



# 课程内容概述

- Python编程技术
- 数据科学基础原理与技术
- 机器学习基础
- 图像处理、时间序列分析基础
- 神经网络与深度学习基础



# Python语言部分

- 容器-迭代器、生成器 ★ ★ ★
- 函数、匿名函数与列表生成式、函数式编程 ★ ★ ★
- 类定义-实例化、可执行类、迭代器类 ★ ★ ★
- 类继承、装饰器类、设计模式简介 ★ ★
- 进程、线程 ★
- 协程 ★ ★ ★, 异步编程实现 ★ ★
- TCP/IP Socket,网络编程 ★
- Tkinter\*



# 关于python:

- 是一个基于词典架构的，面向列表数据加工的语言
- 函数闭包与类继承机制实现了模块化编程方案



# Numpy与线性代数-线性空间

- 矩阵基本操作：定义、计算、维度变换、矩阵切片 ★ ★ ★
- 矩阵运算、广播机制、向量运算 ★ ★ ★
- 矩阵特征值、特征向量；矩阵分解 ★ ★ ★



# 基于矩阵计算的数据科学原理与技术

- 矩阵计算及相关物理意义 ★ ★ ★
- 矩阵分解与特征降维、降噪、协同过滤 ★ ★ ★
- 向量空间-距离-聚类 ★ ★ ★



# Pandas数据处理

- 表数据处理基础、apply函数使用、缺失数据处理 ★ ★ ★
- 聚合操作与统计分析、表连接-融合 ★ ★ ★
- 高维表处理，透视表变换 ★



# Python与机器学习

- 贝叶斯分类器 ★ ★ ★
- 线性回归、logistic regression ★ ★ ★
- K-means、KNN ★ ★ ★
- 图分析与SNA基础 ★
- 基于关联的推荐模型，HITS算法 ★ ★
- HMM序列（线性）模型 ★





# 关于自编码学习

- 聚类
- 回归
- 矩阵分解



# 图像处理、时间序列分析基础

- 图像编码与基本概念 ★ ★
- 图像特征与图像建模 ★ ★
- DCT与图像压缩 ★
- 时间序列处理基础：趋势-周期 消除 ★ ★
- 序列回归、平滑滤波 ★
- 平稳序列概念 ★
- 序列特征挖掘与预测 ★



# 神经网络与深度学习基础

- 神经网络原理与模型实现 ★ ★ ★
- 深度学习架构方案与实现 ★
- Pytorch张量计算 ★ ★ 梯度回传机制 ★
- CNN、RNN-LSTM ★ ★ ★
- Transformer建模与预训练模型基础 ★



# 关于大语言模型

- 基于attention的序列模型
- 多层次多角度话题模型



# 笔试内容分布:

- Python语言技术 (30-35%)
- Numpy、pandas及特征分析、数据分析 (25-30%)
- CV、TSA、图分析 (10-15%)
- 机器学习概念与方法 (10-15%)
- 神经网络、深度学习及建模 (15-20%)



# 与往年相比的新特点：

- 考虑增加部分模型分析的题目，考察同学对不同模型方案在某个具体问题应用上效果差异的理解与改进思路。



# 课程总体评价：

- 课程转型顺利开展
- 大作业环境基本成型
- 为明年开启高水平教学课程提供了基础



# 期末考试时间-地点：

- 考试时间为：2024/06/16 星期日
- 考试时间：下午，14:00 – 16:00
- 考试地点：二教107
- 考试内容：主要是课上讲的基础内容。重点考察作业涉及的知识和技术。
- 考试形式：闭卷笔试





## 关于缓考

- 学生缓考要在提交缓考申请单，**一律由学院审批；**  
学院批准的缓考，会要求学生把缓考回执单交给任课教师。
- 课程开考后递交的缓考申请和病假条无效。
- 任课教师在考场记录上标注该生缓考，成绩系统中的缓考由教务老师录入。

感谢助教团队的辛勤工作！

感谢同学的积极参与与合作！

预祝大家取得好成绩！

