一、课程总览

----------------------------------

1、人工智能包含机器学习，机器学习包含深度学习

2、有监督机器学习流程

历史数据 -> 代入算法，进行模型训练 -> 模型（最优解） -> 对于新数据做出预测

3、有监督机器学习训练流程

---》针对海量数据，进行模型调参的过程

4、sklearn库介绍

---》数学计算包：numpy、scipy、pandas

---》sklearn：简单有效的工具对于数据挖掘，基于numpy，scipy，matplotlib，做了一些算法的实现

---》scipy：傅里叶变换

---》matplotlib：画图库

5、sklearn库用途

（1）分类：识别分类，类别本身存在，如：垃圾邮件识别、图像识别

（2）回归：预测未来程度和趋势，如：药物反应、股票价格

（3）聚类：无监督机器学习，自行对数据划分类型，如：划分客户群体、分组实验结果

（4）降维：（即忽略无关和次要因素）当数据量非常大或需要可视化时，需要降低维度；

---》计算维度越多，预测更准确；但是计算量大，计算时间长；但因为数据量越大模型越准确，所以只能通过降维来加快计算速度

（5）模型选择：同种或不同算法都可能算出不同模型，进行比较、验证、评估，调优选择超参数和最优模型

---》超参数：如学习率、迭代次数等

（6）数据预处理：数据转换（文本转数字）、特征提取（高矮胖瘦）、归一化等

6、线性回归

---》如果模型满足两个假设（线性和回归）

---》线性假设：呈现线性变化；

---》回归假设：服从正态分布。

7、人工智能应用方向

（1）推荐系统

（2）图像识别

（3）自然语言处理（NLP）：情感分析、文本分类、智能客服

（4）其他，如：股票预测、房价预测、论文查重等

=================================================

二、机器学习

1、什么是机器学习

---》让机器像人具备学习的能力（拟人）

2、人 与 机器 类比

---》知识 -- 数据

---》算法 -- 公式

---》神经网络 -- 各种算法，人来帮助机器选择算法

---》模型 -- 参数

---》预测 -- 把新的数据和参数进行计算得到结果

3、机器学习过程简化

---》机器学习目标： 求得类似的公式：y = a\*x + b

---》数据： y1，x1和y2，x2事实上就可以作为数据

---》因素和结果： x为影响结果的因素，y为结果

---》超参数： a，b就是要求解的参数

---》预测： 就是当新的x带到模型公式里面，求得y的预测结果

4、机器学习分类

---》有监督的机器学习：公式中有y

---》无监督的机器学习：公式中没有y

5、有监督的机器学习

---》回归：预测结果具体的值（y是连续的）

---》分类：预测结果是哪个类别（y是离散的），如：输赢

6、无监督的机器学习

---》聚类：在没有y的情况下，把数据分类成一堆堆的

---》降维：X1...Xn维度多，n特别大的时候，需要为了更快的解方程组，减少X数量的算法

7、机器学习、数据挖掘、数据分析

---》机器学习：主要研究算法本身

---》数据挖掘：会用机器学习的算法和数据库的技术

---》数据分析：针对表的操作，如：excel

8、框架列举

numpy 数学计算框架

pandas 数据分析框架

scipy 物理计算框架

matplotlib 绘图的框架

scikit-learn 机器学习的框架

tensorflow 谷歌开源出来的深度学习的框架

keras 开源出来的深度学习的框架

Spark、MLlib 大数据处理开源框架