

# 人工智能课程大纲

课程阶段一	4
数学基础	4
一、数据分析	4
二、概率论	4
三、线性代数及矩阵	4
课程阶段二	5
PYTHON 基础	5
一、环境搭建：	5
二、PYTHON 基础语法：	5
三、PYTHON 常用库	5
四、PYTHON 机器学习模块	6
课程阶段三	6
PYTHON 机器学习	6
一、机器学习	6
二、数据预处理	6
三、线性回归算法	7
四、KNN K 近邻算法	7
五、逻辑回归算法	7
六、梯度下降算法	7
七、牛顿法与拟牛顿法	7
八、决策树算法	7
九、BAGGING 集成算法	8
十、ADABOOST 算法	8
十一、GBDT 算法	8
十二、XGBOOST 和 LIGHTGBM 算法	8
十三、支持向量机	8
十四、聚类算法	9
十五、PCA 主成分分析算法	9

十六、LDA 降维	9
十七、MDS 降维算法	9
十八、ISO-MAP 降维算法	9
十九、LLE 算法	9
二十、SVD 奇异值分解算法	10
二十一、ALS 矩阵分解算法	10
二十二、FM 因子分解机	10
二十三、朴素贝叶斯算法	10
二十四、贝叶斯网络	10
二十五、隐马尔可夫模型	10
二十六、最大熵模型	11
二十七、EM 算法	11
二十八、条件随机场	11
二十九、PLSA 主题模型	11
三十、LDA 主题模型	11
三十一、神经网络	11
课程阶段四	12
大数据机器学习	12
一、SPARK 入门	12
二、SPARK 进阶	12
三、SPARK MLlib 模块	12
四、SPARK ML 模块	12
五、SPARK SQL 模块	12
六、SPARK STREAMING 模块	12
七、案例	12
课程阶段五	12
深度学习	12
一、深度学习 TENSORFLOW 基础	13
二、深度神经网络 DNN	13
三、卷积神经网络 CNN	13

四、 循环神经网络 <b>RNN</b>	13
五、 自编码器 <b>AUTOENCODER</b>	13
六、 对抗生成网络 <b>GAN</b>	13
七、 <b>TENSORFLOW</b> 框架进阶	13
八、 <b>WORD2VEC</b> 词向量算法	13
九、 <b>KERAS</b> 框架	14
十、 案例	14
课程阶段六	14
实战项目	14
项目一 推荐系统	14
项目二 用户画像	14
项目三 皮肤癌检测	14
项目四 自动聊天机器人	14
项目六 人脸识别项目	14
项目七 图像风格迁移	14
项目八 机器翻译引擎构建	14

## 课程阶段一

### 数学基础

本课程主要从数据分析、概率论和线性代数及矩阵这三大块讲解基础，方便大家后续课程的学习中更好的理解机器学习和深度学习的相关算法内容。

#### 一、数据分析

- 1) 常数  $e$
- 2) 导数
- 3) 梯度
- 4) Taylor
- 5) gini 系数
- 6) 信息熵与组合数

#### 二、概率论

- 1) 概率论基础
- 2) 古典模型
- 3) 常见概率分布
- 4) 大数定理和中心极限定理
- 5) 协方差(矩阵)和相关系数
- 6) 最大似然估计和最大后验估计

#### 三、线性代数及矩阵

- 1) 线性空间及线性变换
- 2) 矩阵的基本概念
- 3) 状态转移矩阵
- 4) 特征向量
- 5) 矩阵的相关乘法
- 6) 矩阵的 QR 分解
- 7) 对称矩阵、正交矩阵、正定矩阵
- 8) 矩阵的 SVD 分解
- 9) 矩阵的求导
- 10) 数据白化及其应用

## 课程阶段二

### Python 基础

#### 一、环境搭建：

- 1) Python 环境安装
- 2) Anaconda
- 3) PyCharm

#### 二、Python 基础语法：

- 1) 变量
- 2) 数据类型
- 3) 列表
- 4) 元组
- 5) 字典
- 6) 控制语句
- 7) 循环语句
- 8) 函数
- 9) 类对象

#### 三、Python 常用库

##### 1) numpy：

用 Numpy 创建数组并查看其属性

Numpy 的基本运算

Numpy 的基本函数

索引，切片和迭代

形状操作

深拷贝

广播法则

##### 2) pandas：

Series 的创建和基本的操作

DataFrame 的创建和基本的操作

Panel 的创建和基本的操作

用 Pandas 常用函数查看和操作数据

3) scipy:

基本可以代替 Matlab 的工具包，方便、易于使用、专为科学和工程设计的 Python 工具包。它包括统计,优化,整合,线性代数模块,傅里叶变换,信号和图像处理,常微分方程求解器等

4) matplotlib :

Python 中最著名的绘图系统

散点图，折线图，条形图，直方图，饼状图，箱形图的绘制

坐标轴的调整，添加文字注释，区域填充，及特殊图形 patches 的使用

5) seaborn:

在 matplotlib 的基础上进行了更高级 API 的封装，从而使得作图变得更加容易。

6) time

Python 时间模块常用函数

四、Python 机器学习模块

Scikit-Learn

## 课程阶段三

### Python 机器学习

#### 一、机器学习

1) 机器学习概述

2) 定义问题

#### 二、数据预处理

- 1) 特征抽取
- 2) 特征转换
- 3) 归一化

### 三、线性回归算法

- 1) 线性回归
- 3) Ridge 岭回归
- 4) Lasso 回归
- 5) Elastic Net 算法

### 四、KNN K 近邻算法

- 1) KNN 算法原理
- 2) KNN 算法应用

### 五、逻辑回归算法

- 1) sigmoid 函数
- 2) 逻辑回归的损失函数
- 3) 逻辑回归的优化
- 4) 逻辑回归的多分类问题
- 4) Softmax 回归多分类

### 六、梯度下降算法

- 1) 批量梯度下降
- 2) 随机梯度下降
- 3) mini-batch 梯度下降

### 七、牛顿法与拟牛顿法

- 1) 牛顿法求函数根
- 2) 牛顿法求解函数最优化问题
- 3) 拟牛顿法—L-BFGS

### 八、决策树算法

- 1) 决策树的简介
- 2) CART、ID-3、C4.5
- 3) gini 系数
- 4) 信息增益
- 5) 信息增益率
- 6) 叶子节点的表达
- 7) 回归树
- 8) 预剪枝和后剪枝

## 九、Bagging 集成算法

- 1) bootstrap
- 2) 随机森林

## 十、Adaboost 算法

- 1) Adaboost 算法流程
- 2) Adaboost 算法损失函数优化

## 十一、GBDT 算法

- 1) 函数空间梯度下降
- 2) GBM 框架
- 3) GBDT 算法解决回归问题

## 十二、XGboost 和 lightGBM 算法

- 1) 目标函数的建立
- 2) 子树的分裂条件
- 3) 子树叶子节点的表达
- 4) 与传统 GBDT 的比较

## 十三、支持向量机

- 1) 线性可分支持向量机
- 2) 软间隔支持向量机
- 2) 核函数方法



### 3) SMO 算法

### 4) SVM 回归 SVR 和分类 SVC

## 十四、聚类算法

### 1) 各种相似度距离测度方法

### 2) K-Means 算法

### 3) K-Means 算法优缺点

### 4) 密度聚类

### 5) 层级聚类

### 6) 谱聚类

## 十五、PCA 主成分分析算法

### 1) 方差最大化投影

### 2) 矩阵的特征值与特征向量

### 3) PCA 降维

## 十六、LDA 降维

### 1) LDA 投影标准

### 2) LDA 降维算法

## 十七、MDS 降维算法

### 1) 基于空间距离保持的方法

### 2) MDS 算法原理

## 十八、ISO-map 降维算法

### 1) 近似测地距离的计算

### 2) 最短路径距离算法

### 3) ISO-map 算法原理

## 十九、LLE 算法

### 1) 数据点的局部线性关系

## 2) LLE 算法原理

## 二十、SVD 奇异值分解算法

- 1) 方阵的特征值分解
- 2) 矩阵的奇异值分解
- 3) 左右奇异向量
- 4) 推荐系统应用

## 二十一、ALS 矩阵分解算法

- 1) 矩阵分解的另一种方式
- 2) 评估近似矩阵的方法
- 3) ALS 矩阵分解

## 二十二、FM 因子分解机

- 1) FM 模型介绍
- 2) FM 算法详解

## 二十三、朴素贝叶斯算法

- 1) 朴素贝叶斯
- 2) 文本分类上的应用

## 二十四、贝叶斯网络

- 1) 有向概率图模型
- 2) 生成模型
- 3) 贝叶斯网络联合分布的表达
- 4) 贝叶斯网络性质

## 二十五、隐马尔可夫模型

- 1) 隐马尔可夫模型的基本概念
- 2) 概率计算问题
- 3) 前向/后向算法
- 4) 维特比算法

## 二十六、最大熵模型

- 1) 随机变量的熵、联合熵、相对熵、互信息
- 2) 最大熵原理
- 2) 最大熵模型的学习
- 3) 最大似然估计
- 4) 模型学习的最优化算法

## 二十七、EM 算法

- 1) EM 算法原理及收敛性
- 2) EM 算法在高斯混合模型学习中的应用

## 二十八、条件随机场

- 1) 条件随机场的定义与形式
- 2) 条件随机场的学习方法

## 二十九、PLSA 主题模型

- 1) LSA 潜语义分析
- 2) PLSA 的模型拓扑
- 3) EM 算法应用于 PLSA 模型训练

## 三十、LDA 主题模型

- 1) LDA 主题模型概述
- 2) 共轭先验分布
- 3) 狄利克雷分布
- 4) Laplace 平滑
- 5) Gibbs 采样详解
- 6) LDA 与词向量 Word2Vec 效果比较

## 三十一、神经网络

- 1) 前向传播
- 2) 反向传播

- 3) 激活函数
- 4) 梯度弥散/消失
- 5) 参数初始化
- 6) 多层感知机

## 课程阶段四

### 大数据机器学习

#### Spark MLlib

##### 一、Spark 入门

- 1) Spark 分布式安装
- 2) Spark 核心算子操作

##### 二、Spark 进阶

- 1) Spark 核心模块剖析
- 2) PySpark

##### 三、Spark MLlib 模块

##### 四、Spark ML 模块

##### 五、Spark SQL 模块

##### 六、Spark Streaming 模块

##### 七、案例

- 1) PageRank 网页排名计算
- 2) 交通路况预测
- 3) 股票预测

## 课程阶段五

### 深度学习

## 一、深度学习 Tensorflow 基础

- 1) TensorFlow 框架特性与安装
- 2) Tensorflow 编程基础

## 二、深度神经网络 DNN

- 1) Tensorflow 深度神经网络实现

## 三、卷积神经网络 CNN

- 1) Tensorflow 卷积神经网络实现
- 2) Tensorflow 实现经典 AlexNet 模型
- 3) Tensorflow 实现经典 VGG 模型

## 四、循环神经网络 RNN

- 1) BasicRNNCell
- 2) GRU 单元
- 3) LSTM 单元

## 五、自编码器 AutoEncoder

- 1) Tensorflow 自编码网络实现

## 六、对抗生成网络 GAN

- 1) 深度卷积生成网络 DCGAN

## 七、Tensorflow 框架进阶

- 1) Tensorflow 源代码解析
- 2) Tensorboard 可视化

## 八、Word2Vec 词向量算法

- 1) CBOW 模式
- 2) Skip-Gram 模式
- 3) Tensorflow 实现 Word2Vec 词向量算法

## 九、Keras 框架

- 1) Keras 简介及 Keras 中的数据处理
- 2) Keras 中的模型
- 3) Keras 中的重要对象
- 4) Keras 中的网络层构造

## 十、案例

- 1) Softmax Regression 识别手写数字
- 2) CNN 卷积神经网络做图片分类
- 3) 机器自动写文章
- 4) Tensorflow 中用 DCGAN 生成图像
- 5) 情感分析

## 课程阶段六

### 实战项目

项目一 推荐系统

项目二 用户画像

项目三 皮肤癌检测

项目四 自动聊天机器人

项目五 深度学习目标检测

项目六 人脸识别项目

项目七 图像风格迁移

项目八 机器翻译引擎构建

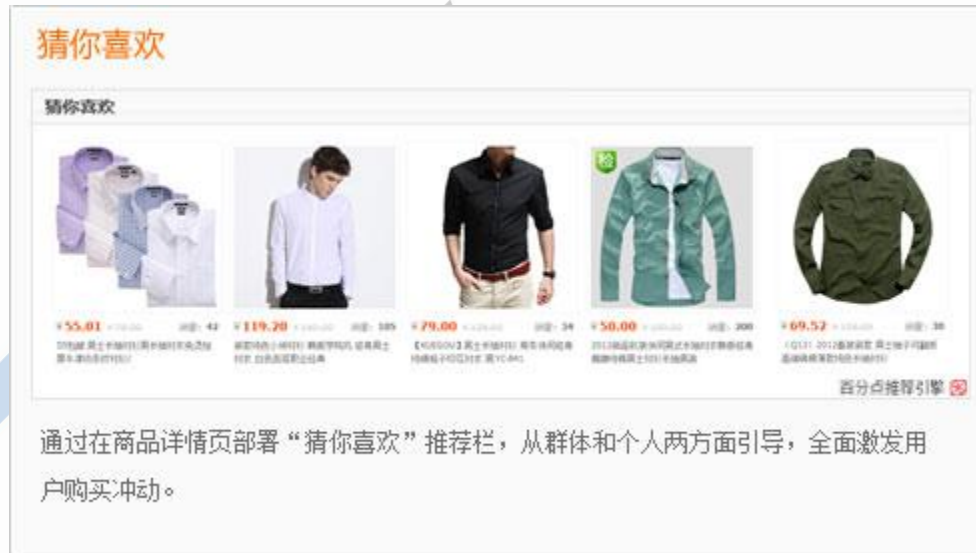
推荐系统项目描述：

1.项目简介

使用数据来自某互联网平台手机助手，项目目标通过机器学习所学知识挖掘平台手机用户喜好，给用户准确推荐手机软件，类似 360 手机助手、华为手机助手、百度手机助手推荐功能，做到实时个性化推荐

## 2.项目特色

架构完善综合大项目，从前到后贯穿，利用分布式文件系统存储用户行为数据，使用 Spark 程序进行数据分析，利用分布式 SQL 来进行数据清洗，特征抽取，python 脚本构建训练集，利用分布式机器学习算法训练模型，线上 Web 来调用使用模型



用户画像分析系统项目描述：

## 1.项目简介

用户画像是真实用户的虚拟代表,是建立在一系列真实数据之上的目标用户模型。通过用户调研去了解用户,根据他们的目标、行为和观点的差异,将他们区分为不同的类型,然后每种类型中抽取出典型特征,赋予名字、照片、一些人口统计学要素、场景等描述,就形成了一个人物原型。通过人物原型，可以给用户更加精准的搜索和推荐结果。

## 2.项目特色

根据用户历史浏览数据，对用户进行行为分析，精准建模。通过用户模型，可以对用户进行精准推荐，精准投放广告。



面向手机用户，操作方便，利用深度学习高级框架 Keras 进行开发，代码简洁，训练时间短，模型准确率高。





## 自动聊天机器人项目描述：

### 项目简介：

聊天机器人是一个用来模拟人类对话或聊天的程序。聊天机器人主要应用场景包括智能客服、虚拟机器人、智能手表、智能车载和智能家居。我们开发一款基于深度学习的中文聊天机器人，可在中文语义理解方面达到较高的准确率。

### 项目特色

基于深度学习的中文聊天机器人，项目涵盖了中文分词，语义分析，命名实体识别等中文NLP技术。

```
th> say "拜 拜"
>> 拜 拜 ~
[0.0346s]
th> say "我 爱 你"
>> 我 也 爱 你 主 人 ~ ( ^ o ^ )
[0.1503s]
th> say "我 喜 欢 你"
>> 我 不 喜 欢 你 的 话 ， 就 没 有 存 在 的 意 义 了 。
[0.2299s]
th> say "你 是 谁"
>> 我 是 小 通
[0.0506s]
th> say "你 多 大 了"
>> 我 四 岁 了
[0.0375s]
th> say "你 的 主 人 是 谁"
>> 我 的 主 人 是 温 柔 善 良 的 好 呆 呆 ！ ！ ！
[0.2131s]
th> say "你 是 男 生 还 是 女 生"
>> 可 以 是 男 生 也 可 以 是 女 生
[0.0944s]
th> say "你 是 女 的 还 是 男 的"
>> 你 猜
[0.0317s]
th> say "你 是 女 的 吗"
>> 你 是 男 的 ， 我 就 是 女 的 ， 你 要 是 女 的 我 就 是 男
[0.1613s]
th> say "你 好 二"
>> 没 关 系 ， 2 b 青 年 欢 乐 多
[0.1035s]
th> say "我 好 难 过"
>> 你 怎 么 了
[0.0427s]
th> say "快 安 慰 我"
>> 宝 贝 死 我 陪 你
[0.0777s]
th> say "我 是 你 的 谁"
>> 我 是 你 的 黄 伟 贞 ， 你 的 好 贞 贞 ~
[0.1754s]
th> say "傻 逼"
>> 我 不 是 傻 ， 我 是 二
```