

# 百面机器学习作业（第三期）

笔记整理人：天国之影

## 说明

1. 每周三、周六为休息日，当天无须打卡，不会安排任何作业和任务。若学习时长中包含周三或周六，则默认忽略当天计划。
2. 本课程作业的所有代码都要基于 Python3，在 Jupyter Notebook 上完成。
3. 课程老师：Jaster 老师

我的作业 GitHub 地址（在每一个 Week 中均有一个 MyHomeWork 文件夹，用于记录我的作业完成情况，所有 ipynb 文件均带注释）：

## 1 第 1 周

### 1.1 了解机器学习中特征工程和模型评估

#### 前言：

人工智能技术正在对社会结构、职场、教育等带来革命性的变化。未来几年是人工智能技术全面普及化的时期，也是该技术的相关人才最为稀缺的时期。我们希望通过面试刷题班，帮助对人工智能和机器学习感兴趣的朋友更加深入地了解这个领域的基本技能，帮助已经有计算机技术基础的同行们，成为 AI 领域的 offer 收割机。

**知识要求：**（学习的过程中可以遇到问题后再复习）

- （1）了解高等数学、概率论、线性代数基础知识。
- （2）了解基础机器学习算法，若基础机器学习算法不熟悉建议可参考西瓜书进行适当复习。

#### 资料下载：

《百面机器学习：算法工程师带你去面试》

电子版链接：<https://share.weiyun.com/5k9AYXR>

密码: [swqwd9](#)

### 本日任务简介:

- (1) 学习绪论视频, 了解预备知识
- (2) 学习《百面机器学习》1.1-1.7, 2.1-2.7

### 详细说明:

- (1) 观看绪论视频
- (2) 学习《百面机器学习》1.1-1.7, 2.1-2.7

今天学习任务比较简单, 因此只有书籍阅读任务, 无补充图文或视频教程。之后的难点任务, 均会有老师录制成讲解视频。并且每周六老师会对当周疑难问题进行直播讲解。

本节将会向大家介绍特征工程和模型评估当中的一些知识点。特征工程是对原始数据进行一系列工程处理, 将其提炼为特征, 作为输入供算法和模型使用。做过项目或者竞赛的都应当了解特征工程的重要性, 面对不在一个数量级的特征, 类别性特征, 高维特征, 特征组合等等场景, 我们应该怎么做。除此之外, 我们还需要对自己领域内, 比如自然语言处理, 图像处理等有哪些特征工程的技巧有所了解。

模型评估是构建一个模型重要的一环, 分类问题、排序问题、回归问题往往需要使用不同的指标进行评估。在这里我们将着重攻克常见的指标, 如精确率, 召回率, F1 值, P-R 曲线, ROC 曲线和 AUC。通过指标我们观察到欠拟合和过拟合时, 该如何处理。

### 打卡要求:

说明对于不同的特征该如何进行特征工程? 模型评估中不同的指标用在什么场景中?

打卡内容: 不少于 150 字

打卡截止提交时间: 8/21

## 1.2 LeetCode 刷题

学习时长: 8/21—8/25

### 任务简介:

LeetCode 算法作业打卡日，完成本周 LeetCode 题目作业，并且提交通过，进行打卡

**详细说明：**答案在下周 LeetCode 任务布置后公布

### **Leetcode 算法题**

(1) 双指针（题号：167）：

<https://leetcode.com/problems/two-sum-ii-input-array-is-sorted/description/>

排序：

(2) 快速选择、堆排序（题号：215）：

<https://leetcode.com/problems/kth-largest-element-in-an-array/description/>

(3) 桶排序（题号：347）：

<https://leetcode.com/problems/top-k-frequent-elements/description/>

(4) 荷兰国旗问题（题号：75）：

<https://leetcode.com/problems/sort-colors/description/>

(5) 贪心（题号：455）：

<https://leetcode.com/problems/assign-cookies/description/>

（本次 leetcode 中涉及到排序的问题请同学们不要调用系统库函数去实现，尝试自己手动实现。）

**打卡要求：**LeetCode 作业中所有题目通过截图提交

打卡内容：至少上传 6 张图片

打卡截止提交时间：8/25

## **1.3 了解监督学习中的经典算法**

学习时长：8/22—8/23

**任务简介：** 学习《百面机器学习》3.2-3.3

**详细说明：**

由于支持向量机部分较难，我们先行学习后面两章。逻辑回归、决策树这两种经典分类算法。在很多实际问题中，因为海量且带有精确标注的数据难以获得，深度学习没有大显身手的余地，许多传统方法可以灵活巧妙地进行处理。

我们为大家准备了逻辑回归和决策树的讲解视频和算法代码实战，视频和书本搭配可以更好的理解知识点，大家可以边看讲解文档，边运行程序代码，以期能够更好的理解这两个算法。

**学习视频：**

逻辑回归

决策树

**实战代码：**

链接：<https://pan.baidu.com/s/1wa73n4LshEiylpRHViUMbA>

密码：[bbwl](#)

在同学们完成了第一部分的代码实战后，我们准备了一个达观杯的文本智能处理挑战赛，并且使用逻辑回归进行分类，很简单的一个版本，让大家体验一下比赛的感觉，具体的可以打开下面的链接，下载相应的文档，进行学习。

链接：<https://pan.baidu.com/s/1MUAYdLfqEgqp0tUkH4iiHQ>

密码：[noqn](#)

**打卡要求：** 描述两类算法的应用场景、优势和区别

**打卡内容：** 不少于 100 字描述

**打卡截止日期：** 8/23

## 1.4 本周学习内容总结

**任务简介：** 第一周 LeetCode 作业

**详细说明：** 请一定在今天完成本周 LeetCode 全部题目

**打卡要求：** 请用文字描述，本周所学知识的重点，格式不限，思维导图、手写、

电子版均可

打卡内容：不少于 50 字

打卡截止日期：8/24

## 2 第 2 周

### 2.1 学习支持向量机

学习时长：8/26—8/27

任务简介：学习《百面机器学习》3.1

详细说明：

在算法面试中，手推 SVM 已经越来越像快速排序一样，成为了一道必点菜。而 SVM 的原理的推导实际上需要一定的数学基础，比如拉格朗日乘子法、核函数等。而越往后学习越能发现，这些数学的基础知识在很多地方都会用到，故而学习 SVM 的原理以及不仅仅是对 SVM 这个算法本身的学习了，它更像是一种基础。我们建议每位同学在学完本章后，都应当能够在一张白纸上推导出 SVM 的算法全过程。在本章我们也提供了一些 SVM 推导的小视频，以供大家学习。

在下面的百度云链接里，我们提供了视频中的文稿以及实战部分的代码，希望大家不惜一切代价，学会 SVM。

链接：<https://pan.baidu.com/s/1nH4TbKaXFFTKhU6N3AQ3Tw>

提取码：[tg6u](#)

视频：

- (1) 几个重要的概念
- (2) SVM 最优化问题
- (3) 硬间隔 SVM 最优化问题的推导
- (4) 线性可分 SVM
- (5) 核函数
- (6) SMO 算法

打卡要求：手动将 SVM 的数学原理推导一遍，并拍照提交

打卡内容：至少上传 1 张图片

打卡截止日期：8/27

## 2.2 LeetCode 刷题

学习时长：8/27—9/1

任务简介：

LeetCode 算法作业打卡日，完成本周 LeetCode 题目作业，并且提交通过，进行打卡

详细说明：答案在下周 LeetCode 任务布置后公布

**Leetcode 算法题：**

二分（题号：69）：

<https://leetcode.com/problems/sqrtx/description/>

分治（题号：241）：

<https://leetcode.com/problems/different-ways-to-add-parentheses/description/>

链表（题号：160）：

<https://leetcode.com/problems/intersection-of-two-linked-lists/description/>

哈希表（题号：1）：

<https://leetcode.com/problems/two-sum/description/>

字符串（题号：242）：

<https://leetcode.com/problems/valid-anagram/description/>

栈和队列（题号：232）：

<https://leetcode.com/problems/implement-queue-using-stacks/description/>

第一周的 LeetCode 参考答案来了，同学们可以看答案和讲解视频对比来看一下。

链接: <https://pan.baidu.com/s/1QW2l4qx3t1YNINEKLyv4VA>

提取码: 6tod

**LeetCode 解题视频:**

双指针(题号: 167)

快速选择、堆排序(题号: 215)

桶排序(题号: 347)

荷兰国旗问题(题号: 75)

贪心(题号: 455)

**打卡要求:** LeetCode 作业中所有题目通过截图提交

打卡内容: 至少上传 6 张图片

打卡截止提交时间: 9/1

## 2.3 了解机器学习中如何降维处理

**任务简介:** 学习《百面机器学习》4.1-4.4

**详细说明:**

不管是机器学习还是深度学习,数据都是以向量的形式进行表示的。而对高维向量进行处理时会极大的消耗系统资源,甚至产生维度灾难。因而降维不仅是在面试中,同时在实际应用中都是非常重要的知识点。本章主要介绍了 PCA 和 LDA 这两种降维的算法,其中 PCA 又从两个角度进行推导,并在最后将 PCA 和 LDA 进行对比,两种相似度较高的算法在一起对比是比较好的一种学习方法。在学习本章的时候我们需要特别注意高维向量如何在低维基向量中进行表示,这一点特别重要。同时我们也希望同学们能够在学习完本章后能够理解这两种算法的推导过程,最好是能自己从头推导出来。

在下面的百度云里,为大家准备了 PCA 的实战代码:

链接: <https://pan.baidu.com/s/1PdW4qzHRe-S6MmkWII7Mbw>

密码: s3s1

我们还为大家录制了两种降维方法 PCA 和 LDA 的推导过程的视频:

PCA 和 LDA

### 打卡要求：

手动在纸上或其他工具进行推导并对每个问题拍照提交

(1) PCA 从两个角度考虑如何进行数学推导？

(2) LDA 的数学原理是什么？

(3) 简述 PCA 和 LDA 的异同？

打卡内容：至少上传 3 张图片

打卡截止日期：8/30

## 2.4 第 2 周直播答疑

直播课件和代码下载

链接：[https://pan.baidu.com/s/1DAD\\_jz0nCmkZW4UN11SyWQ](https://pan.baidu.com/s/1DAD_jz0nCmkZW4UN11SyWQ)

提取码：[swke](#)

## 2.5 本周学习内容总结

学习时长：9/1—9/1

任务简介：第二周 LeetCode 作业截至提示

详细说明：请一定在今天完成本周 LeetCode 全部题目

打卡要求：请用文字描述，本周所学知识的重点，格式不限，思维导图、手写、电子版均可

打卡内容：不少于 50 字

打卡截止日期：9/1

## 3 第 3 周

### 3.1 了解机器学习中的非监督学习算法

学习时长：9/2—9/3

任务简介：学习《百面机器学习》5.1-5.4



### 详细说明:

在日常数据的训练当中往往并没有那么多的标注好的数据,相比于监督学习,非监督学习的输入数据没有标签信息,需要通过算法模型来挖掘数据内在的结构和模式。本章最主要的是要掌握 K 均值算法的步骤、如何进行调优、改进并证明其收敛性。在掌握这个算法的基础上再看 GMM 算法和 SOM 算法。同时还需要重点了解聚类算法是如何评估的,可以和第二章模型评估对比来看有什么异同。

### 视频:

#### K-means

在这里,我们为大家准备了一个 kmeans 算法的标准实战过程,大家再进行 kmeans 聚类的时候,可以通过这样的一个流程进行。

链接: [https://pan.baidu.com/s/1x\\_d-gNN1Vms6CwjN0aTr0w](https://pan.baidu.com/s/1x_d-gNN1Vms6CwjN0aTr0w)

提取码: [kdyr](#)

### 打卡要求:

K 均值算法的步骤、优缺点、如何调优、如何改进、如何证明其收敛性

打卡内容: 不少于 250 字

打卡截止时间: 9/3

## 3.2 LeetCode 刷题

学习时长: 9/3——9/7

任务简介: LeetCode 算法作业打卡日,完成本周 LeetCode 题目作业,并且提交通过,进行打卡

详细说明: 答案在下周 LeetCode 任务布置后公布

### LeetCode 算法题:

(1) 字符串 (题号: 409):

<https://leetcode.com/problems/longest-palindrome/description/>

(2) 数组和矩阵 (题号: 378):

<https://leetcode.com/problems/kth-smallest-element-in-a-sorted-matrix/description/>

(3) 位运算 (题号: 260):

<https://leetcode.com/problems/single-number-iii/description/>

数学

(4) 进制转换 (题号: 504):

<https://leetcode.com/problems/base-7/description/>

(5) 相遇问题 (题号: 462):

<https://leetcode.com/problems/minimum-moves-to-equal-array-elements-ii/description/>

(6) 多数投票问题 (题号: 169):

<https://leetcode.com/problems/majority-element/description/>

**第二周 LeetCode 参考答案:**

链接: <https://pan.baidu.com/s/19qAtmWbc1sXUWFEFb4wH2g>

提取码: [wxc4](#)

**LeetCode 解题视频:**

二分 (题号: 69)

分治 (题号: 241)

链表 (题号: 160)

哈希表 (题号: 1)

字符串 (题号: 242)

栈和队列 (题号: 232)

**打卡要求:** LeetCode 作业中所有题目通过截图提交

打卡内容：至少上传 6 张图片

打卡截止提交时间：9/8

### 3.3 机器学习中的概率图模型

**任务简介：**学习《百面机器学习》6.1-6.5

**详细说明：**

如果用一个词来形容概率图模型（Probabilistic Graphical Model）的话，那就是“优雅”。对于一个实际问题，我们希望能够挖掘隐含在数据中的知识。概率图模型构建了这样一幅图，用观测结点表示观测到的数据，用隐含结点表示潜在的知识，用边来描述知识与数据的相互关系，最后基于这样的关系图获得一个概率分布，非常“优雅”地解决了问题。而本章相比较前面而言也是最难的一个部分。我们希望大家再学习完本章后了解贝叶斯网络和马尔可夫网络的异同、生成式模式和判别式模式的异同，最后还希望大家能够掌握两个非常重要的算法 HMM 和 CRF 算法。

这一节课程非常难，但用到的地方很多，希望大家最起码能够知道概率图模型是个什么东西。本节还提供了一个 hmm 简单分词的实现，学有余力的同学可以参考跑一跑效果。

下载链接：<https://pan.baidu.com/s/1inFK81-zUZQ91EMdxsD0uw>

提取码：[v29p](#)

**视频：**

HMM 的引出和问题的介绍

HMM 预测问题之维特比算法

CRF 的一些基础概念

CRF 具体介绍

**打卡要求：**

- (1) 生成式模式和判别式模式区别是什么？
- (2) 马尔可夫网络和贝叶斯网络的联合概率分布是什么？
- (3) 中文分词问题如何使用 HMM 来进行解决？

打卡内容：不少于 150 字

打卡提交时间：9/6

### 3.4 第 3 周直播答疑

第三次直播答疑 ppt 和代码下载：

链接：<https://pan.baidu.com/s/1RU7Qcs96s20217xo1SJgPg>

提取码：[ldxe](#)

### 3.5 本周学习内容总结

学习时长：9/8——9/8

任务简介：第三周 LeetCode 作业截至提示

本周四的课程非常难，但用到的地方很多，所以老师做了视频讲解和实战代码供大家学习。希望大家最起码能够知道概率图模型是个什么东西。同时为提供了一个 hmm 简单分词的实现，学有余力的同学可以参考跑一跑效果。（返回周四的任务即可查看详情）

链接：<https://pan.baidu.com/s/1inFK81-zUZQ91EMdxsD0uw>

提取码：[v29p](#)

**打卡要求：**请用文字描述，本周所学知识的重点，格式不限，思维导图、手写、电子版均可。

打卡内容：不少于 50 字

打卡截止日期：9/8

## 4 第 4 周

### 4.1 了解优化算法的原理

任务简介：学习《百面机器学习》7.1-7.7

详细说明：

在本节中，我们将会带领大家一起学习优化算法的原理。随着大数据和深度

学习的迅猛发展，在实际应用中面临的大多是大规模、高度非凸的优化问题，这给传统的基于全量数据、凸优化的优化理论带来了巨大的挑战。如何设计适用于新场景的、高效的、准确的优化算法成为近年来的研究热点。优化虽然是一门古老的学科，但是大部分能够用于训练深度神经网络的优化算法都是近几年才被提出，如 Adam 算法等

代码下载链接：

链接：<https://pan.baidu.com/s/1f0un30DQ9e1cdGeo9L9vBQ>

提取码：[3wib](#)

视频：

- (1) 深度学习中的优化问题
- (2) 梯度下降简单的数学原理
- (3) 随机梯度下降和小批量随机梯度下降
- (4) 动量法
- (5) 常见的一些改进的优化算法

打卡要求：

- (1) 有监督学习和无约束优化问题的优化方法分别有哪些？
- (2) 随机梯度下降为什么会失效？

打卡内容：不少于 100 字

打卡截止日期：9/10

## 4.2 LeetCode 刷题

学习时长：9/10—9/15

任务简介：LeetCode 算法作业打卡日，完成本周 LeetCode 题目作业，并且提交通过，进行打卡

详细说明：答案在下周 LeetCode 任务布置后公布

**Leetcode 算法题：**

- (1) 树：

递归（题号：110）：

<https://leetcode.com/problems/balanced-binary-tree/description/>

层次遍历（题号：513）：

<https://leetcode.com/problems/find-bottom-left-tree-value/description/>

前中后序遍历（题号：144）：

<https://leetcode.com/problems/binary-tree-preorder-traversal/description/>

BST（题号：230）：

<https://leetcode.com/problems/kth-smallest-element-in-a-bst/description/>

Trie（题号：208）：

<https://leetcode.com/problems/implement-trie-prefix-tree/description/>

（2）图：

二分图（题号：785）：

<https://leetcode.com/problems/is-graph-bipartite/description/>

拓扑排序（题号：207）：

<https://leetcode.com/problems/course-schedule/description/>

并查集（题号：684）：

<https://leetcode.com/problems/redundant-connection/description/>

**第三周 LeetCode 参考答案：**

链接: <https://pan.baidu.com/s/1kpEu2QAzRQgBOZKav8mGcA>

提取码: [x041](#)

**LeetCode 解题视频:**

- (1) 字符串 (题号: 409)
- (2) 数组和矩阵 (题号: 378)
- (3) 位运算 (题号: 260)
- (4) 进制转换 (题号: 504)
- (5) 相遇问题 (题号: 462)
- (6) 多数投票问题 (题号: 169)

**打卡要求:** LeetCode 作业中所有题目通过截图提交

打卡内容: 至少上传 8 张图片

打卡截止提交时间: 9/14

### 4.3 了解常见的采样方法

学习时长: 9/12——9/13

任务简介: 学习《百面机器学习》8.1-8.7

详细说明:

对于一些简单的分布,如均匀分布、高斯分布等,很多编程语言里面都有直接的采样函数。然而,即使是这些简单分布,其采样过程也并不是显而易见的,仍需要精心设计。对于比较复杂的分布,往往并没有直接的采样函数可供调用,这时就需要其他更加复杂的采样方法。因此,对采样方法的深入理解是很有必要的。

视频:

- (1) 蒙塔卡罗模拟定积分和重要性采样
- (2) 均匀采样、逆变换采样、拒绝采样
- (3) MCMC 采样
- (4) 吉布斯采样

打卡要求:

- (1) 有哪几种采样方法?
- (2) 马尔可夫蒙特卡洛采样法的思想是什么?
- (3) 如何进行不均衡样本的采样?

打卡内容: 不少于 150 字

打卡截止日期: 9/13

## 4.4 本周学习内容总结

学习时长: 9/15—9/15

任务简介: 第四周 LeetCode 作业截至提示

打卡要求: 请用文字描述, 本周所学知识的重点, 格式不限, 思维导图、手写、电子版均可

打卡内容: 不少于 50 字

打卡截止日期: 9/15

## 5 第 5 周

### 5.1 学习前向神经

学习时长: 9/16—9/17

任务简介: 学习《百面机器学习》9.1-9.6

详细说明:

神经网络是我们踏入深度学习的必经之门, 相信大家对神经网络都有所了解或者非常熟悉, 在这里, 我们将复习一下神经网络的基础知识, 包括常见的激活函数优缺点, 重点的反向传播算法, 神经网络的训练技巧, 以及介绍几个深度神经网络等。希望学习过后, 大家对神经网络有进一步的认识

本章和大家一起实现一个简单的浅层神经网络, 在这个过程中, 讲解了前向传播、后向传播这些东西, 希望可以帮助大家理解基础的神经网络这一块  
视频代码下载链接:

链接: <https://pan.baidu.com/s/1bwz04wRtGcHV0jfZyC9dYg>

提取码: o2ar



**视频:**

- (1) 网络图和激活函数
- (2) 前向传播
- (3) 损失函数选用
- (4) 反向传播-1
- (5) 反向传播-2

**打卡要求:**

- (1) 几种激活函数的优缺点?
- (2) 书中提及的神经网络三种训练技巧是什么, 分别解决什么问题?

打卡内容: 不少于 100 字

打卡截止日期: 9/17

## 5.2 LeetCode 刷题

学习时长: 9/17—9/22

**任务简介:** LeetCode 算法作业打卡日, 完成本周 LeetCode 题目作业, 并且提交通过, 进行打卡

**详细说明:** 答案在下周 LeetCode 任务布置后公布

**LeetCode 算法题:**

搜索:

BFS (题号: 279):

<https://leetcode.com/problems/perfect-squares/description/>

DFS (题号: 695):

<https://leetcode.com/problems/max-area-of-island/description/>

Backtracking (题号: 17):

<https://leetcode.com/problems/letter-combinations-of-a-phone-number/description/>

动态规划:

斐波那契数列 (题号: 70):

<https://leetcode.com/problems/climbing-stairs/description/>

矩阵路径 (题号: 64):

<https://leetcode.com/problems/minimum-path-sum/description/>

数组区间 (题号: 303):

<https://leetcode.com/problems/range-sum-query-immutable/description/>

分割整数 (题号: 343):

<https://leetcode.com/problems/integer-break/description/>

最长递增子序列 (题号: 300):

<https://leetcode.com/problems/longest-increasing-subsequence/description/>

0-1 背包 (题号: 416):

<https://leetcode.com/problems/partition-equal-subset-sum/description/>

股票交易 (题号: 309):

<https://leetcode.com/problems/best-time-to-buy-and-sell-stock-with-cooldown/description/>

字符串编辑 (题号: 583):

<https://leetcode.com/problems/delete-operation-for-two-strings/description/>

第四周 LeetCode 刷题参考答案:

链接: <https://pan.baidu.com/s/1BvLJbAWRxCHeejod-sf7iQ>

提取码: [q8uw](#)

**LeetCode 解题视频:**

- (1) 递归 (题号: 110)
- (2) 前中后序遍历 (题号: 144)
- (3) Trie (题号: 208)
- (4) BST (题号: 230)
- (5) 层次遍历 (题号: 513)

**打卡要求:** LeetCode 作业中所有题目通过截图提交

打卡内容: 总共有 11 张图片, 提交 9 张即可

打卡截止提交时间: 9/22

## 5.3 了解序列数据中常用的循环神经网络

学习时长: 9/19——9/20

任务简介: 学习《百面机器学习》10.1-10.6

详细说明:

循环神经网络 (RNN) 是用来建模序列化数据的一种主流深度学习模型, 近年来的应用场景非常的广, 且有了突破性的进展——机器翻译、序列标注、图像描述、推荐系统、智能聊天机器人、自动作词作曲等。在本节, 我们将带领大家探索 RNN 的基本构造, 梯度消失问题, LSTM, Seq2Seq 模型和注意力机制

**视频代码下载链接:**

链接: <https://pan.baidu.com/s/1390SRcHb9hCSVP1E1ULe6w>

提取码: [lkht](#)

**视频:**

- (1) 循环神经网络 RNN
- (2) GRU 和 LSTM

**打卡要求:**

- (1) RNN 为什么会有梯度消失或爆炸, 分别如何解决?
- (2) LSTM 有哪些门, 作用是什么?

(3) Seq2Seq 模型引入注意力机制是为了解决什么问题？

打卡内容：不少于 100 字

打卡截止日期：9/20

## 5.4 本周学习内容总结

学习时长：9/22—9/22

任务简介：第五周 LeetCode 作业截至提示

详细说明：请一定在今天完成本周 LeetCode 全部题目

第五周 LeetCode 答案：

链接：<https://pan.baidu.com/s/18t8TagTwy5R9kU0678ftqw>

提取码：[adpb](#)

**打卡要求：**请用文字描述，本周所学知识的重点，格式不限，思维导图、手写、电子版均可

打卡内容：不少于 50 字

打卡截止日期：9/22