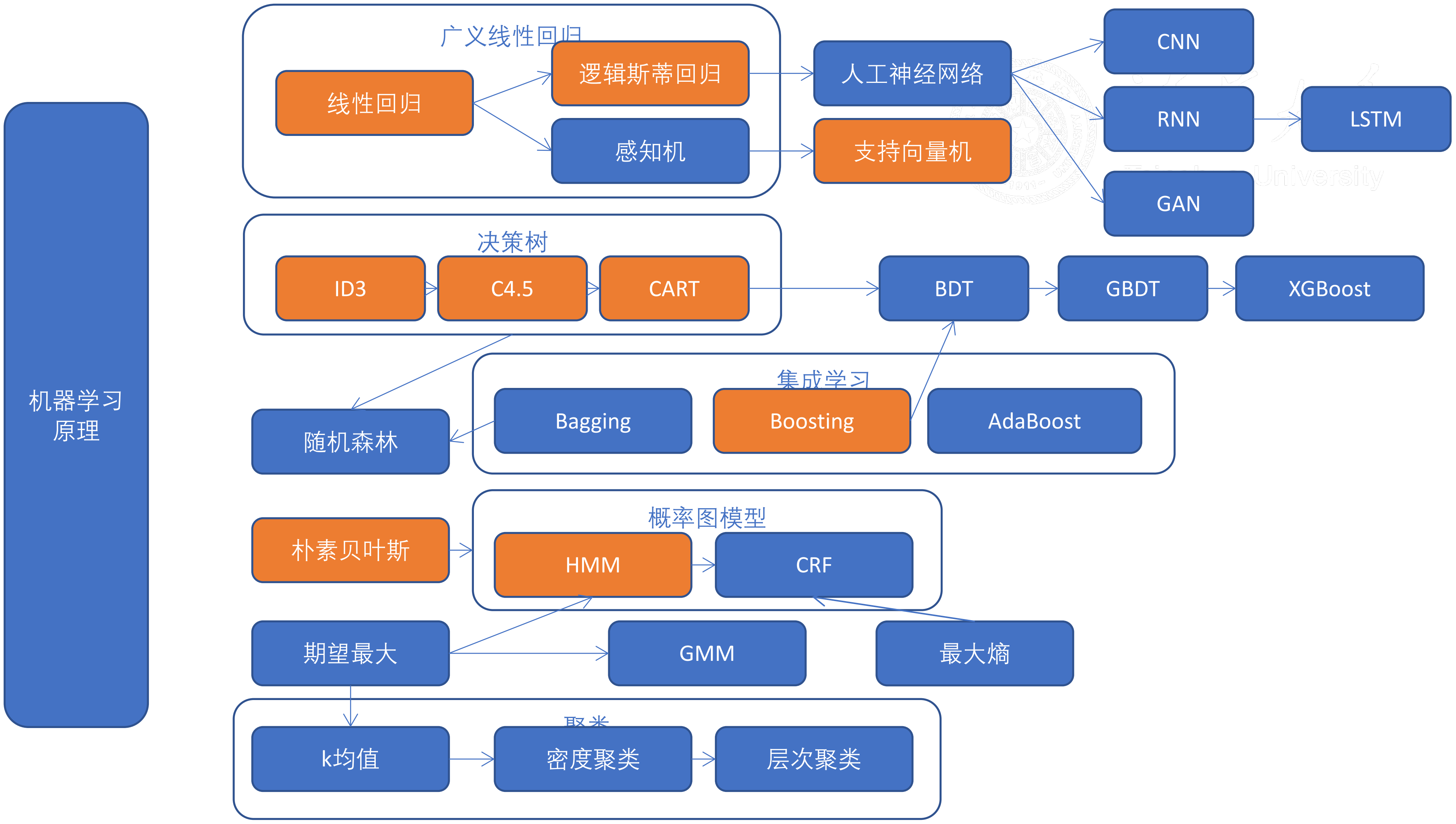

机器学习
原理
广义线性回归
线性回归
逻辑斯蒂回归
感知机
人工神经网络
支持向量机
CNN
RNN
LSTM
GAN
决策树
ID3
C4.5
CART
BDT
GBDT
XGBoost
集成学习
随机森林
Bagging
Boosting
AdaBoost
概率图模型
朴素贝叶斯
HMM
CRF
期望最大
GMM
最大熵
聚类
k均值
密度聚类
层次聚类

第九章
EM期望极大算法

EM算法的导出

- 为什么EM算法能近似实现对观测数据的极大似然估计？
- 极大化(不完全数据)Y关于参数Θ的极大似然函数：

$$L(\theta)=\log P(Y\mid\theta)=\log\sum_Z P(Y,Z\mid\theta)$$
$$=\log\left(\sum_Z P(Y\mid Z,\theta)P(Z\mid\theta)\right)$$

- 难点：有未观测数据，包含和的对数。
- EM通过迭代逐步近似极大化L(Θ),希望 $L(\theta)>L(\theta^{(i)})$

EM算法的导出

$$\log \sum_j \lambda_j y_j \geq \sum_j \lambda_j \log y_j$$

- 考虑二者的差：

$$L(\theta) - L(\theta^{(i)}) = \log\left(\sum_Z P(Y|Z,\theta)P(Z|\theta)\right) - \log P(Y|\theta^{(i)})$$

- Jason不等式：

$$\begin{aligned}
L(\theta) - L(\theta^{(i)}) &= \log\left(\sum_Z P(Y|Z,\theta^{(i)})\frac{P(Y|Z,\theta)P(Z|\theta)}{P(Y|Z,\theta^{(i)})}\right) - \log P(Y|\theta^{(i)}) \\
&\geqslant \sum_Z P(Z|Y,\theta^{(i)}) \log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})} - \log P(Y|\theta^{(i)}) \\
&= \sum_Z P(Z|Y,\theta^{(i)}) \log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})P(Y|\theta^{(i)})}
\end{aligned}$$

EM算法的导出

- 令：$$B(\theta,\theta^{(i)}) \triangleq L(\theta^{(i)}) + \sum_{Z} P(Z|Y,\theta^{(i)}) \log \frac{P(Y|Z,\theta)P(Z|\theta)}{P(Z|Y,\theta^{(i)})P(Y|\theta^{(i)})}$$
- 则：$$L(\theta) \geqslant B(\theta,\theta^{(i)})$$
  $$L(\theta^{(i)}) = B(\theta^{(i)},\theta^{(i)})$$
  任何可以使 $B(\theta,\theta^{(i)})$ 增大的 θ，也可以使 $L(\theta)$ 增大
- 选择：$$\theta^{(i+1)} = \arg\max_{\theta} B(\theta,\theta^{(i)})$$

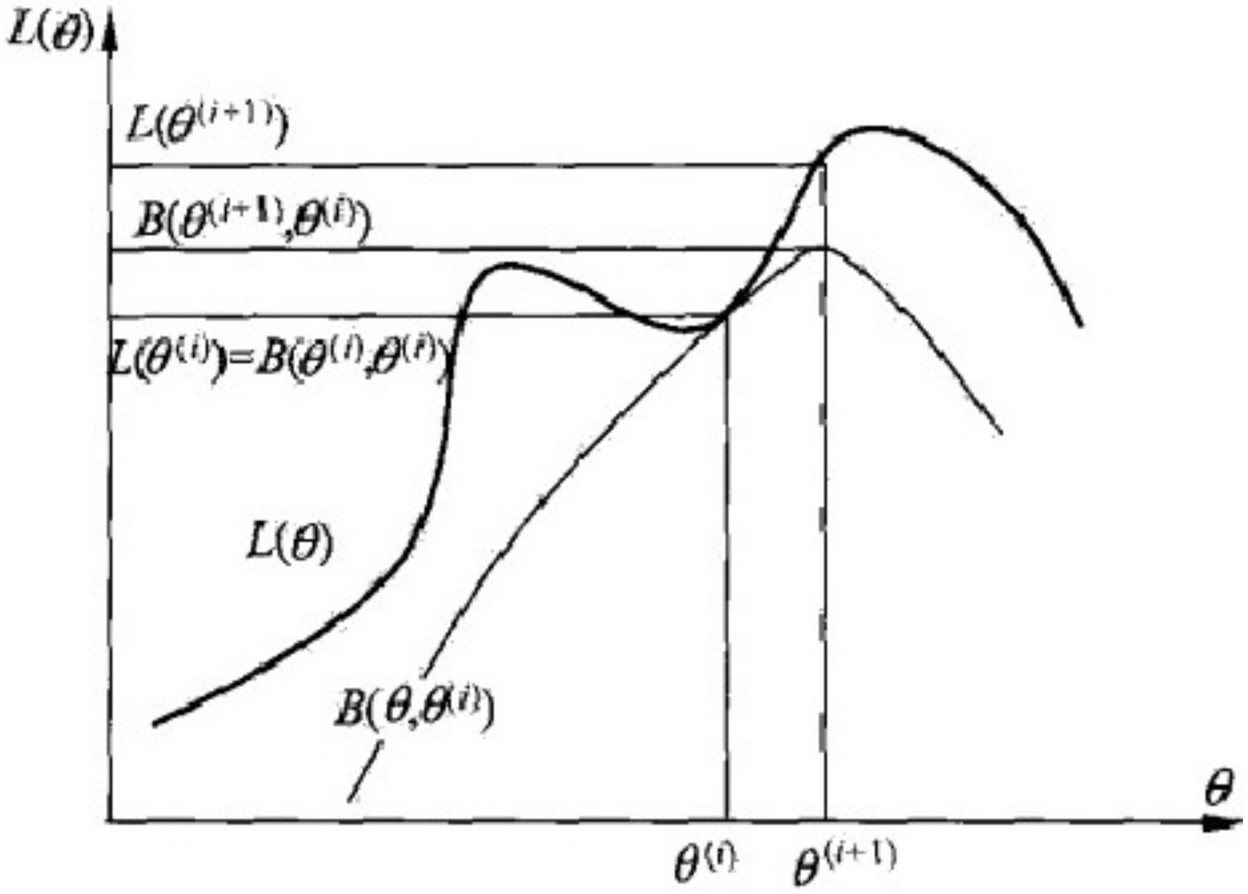
EM算法的导出

- 省去和Θ无关的项：

$$\begin{aligned}
\theta^{(i+1)} &= \arg\max_{\theta}\left(L(\theta^{(i)}) + \sum_{Z} P(Z \mid Y,\theta^{(i)}) \log \frac{P(Y \mid Z,\theta)P(Z \mid \theta)}{P(Z \mid Y,\theta^{(i)})P(Y \mid \theta^{(i)})}\right) \\
&= \arg\max_{\theta}\left(\sum_{Z} P(Z \mid Y,\theta^{(i)}) \log(P(Y \mid Z,\theta)P(Z \mid \theta))\right) \\
&= \arg\max_{\theta}\left(\sum_{Z} P(Z \mid Y,\theta^{(i)}) \log P(Y,Z \mid \theta)\right) \\
&= \arg\max_{\theta} Q(\theta,\theta^{(i)})
\end{aligned}$$

EM算法的解释

L(Θ)开始



EM方法

输入：观测变量数据Y,隐变量数据Z，联合分布P(Y,Z|Θ)

条件分布P(Z|Y, Θ)

输出：模型参数Θ

（1）选择参数的初值 $\theta^{(0)}$，开始迭代；

（2）E 步：记 $\theta^{(i)}$ 为第 i 次迭代参数 θ 的估计值，

在第 $i+1$ 次迭代的 E 步，计算

$$
\begin{aligned}
Q(\theta,\theta^{(i)}) &= E_Z[\log P(Y,Z\mid\theta)\mid Y,\theta^{(i)}] \\
&= \sum_Z \log P(Y,Z\mid\theta)P(Z\mid Y,\theta^{(i)})
\end{aligned}
$$

给定观测数据Y和当前参数估计Θ

EM方法

（3）M步：求使 $Q(\theta,\theta^{(i)})$ 极大化的 θ，

确定第 $i+1$ 次迭代的参数的估计值 $\theta^{(i+1)}$

$$\theta^{(i+1)}=\arg\max_{\theta} Q(\theta,\theta^{(i)})$$

Q函数定义：

完全数据的对数似然函数logP(Y,Z|Θ)关于在给定观测数据Y和当前函数$Θ^{(i)}$下对未观测数据Z的条件概率分布

P(Z|Y, $Θ^{(i)}$),的期望称为Q函数，即：

$$Q(\theta,\theta^{(i)})=E_Z[\log P(Y,Z\mid\theta)\mid Y,\theta^{(i)}]$$

EM方法

- 算法说明：
- 步骤3，完成一次迭代：$\Theta^{(i)}$到$\Theta^{(i+1)}$，将证明每次迭代使似然函数增大或达到局部最大值。
- 步骤4，停止迭代的条件

$$\|\theta^{(i+1)}-\theta^{(i)}\|<\varepsilon_1 \quad \text{或} \quad \|Q(\theta^{(i+1)},\theta^{(i)})-Q(\theta^{(i)},\theta^{(i)})\|<\varepsilon_2$$

EM在非监督学习中的应用

- 生成模型由联合概率分布P(X,Y)表示，可以认为非监督学习训练数据是联合概率分布产生的数据，X为观测数据，Y为未观测数据。

EM算法在高斯混合模型学习中的应用

- 高斯混合模型：
- 概率分布模型；$P(y|\theta)=\sum_{k=1}^{K}\alpha_k\phi(y|\theta_k)$
- 系数：$\alpha_k \geqslant 0$，$\sum_{k=1}^{K}\alpha_k=1$
- 高斯分布密度：$\phi(y|\theta_k)$ $\theta_k=(\mu_k,\sigma_k^2)$
- 第K个分模型：$\phi(y|\theta_k)=\frac{1}{\sqrt{2\pi}\sigma_k}\exp\left(-\frac{(y-\mu_k)^2}{2\sigma_k^2}\right)$

可任意高斯模型

高斯混合模型参数估计的EM算法

- 假设观测数据$y_1,y_2,\ldots.y_N$由高斯混合模型生成：

$$P(y\mid\theta)=\sum_{k=1}^{K}\alpha_k\phi(y\mid\theta_k)$$

$$\theta=(\alpha_1,\alpha_2,\cdots,\alpha_K;\theta_1,\theta_2,\cdots,\theta_K)$$

- 用EM算法估计参数；
- 1、明确隐变量，写出完全数据的对数似然函数：
  - 设想观测数据yi是依概率a_k选择第k个高斯分模型 $\phi(y\mid\theta_k)$ 生成，隐变量

$$\gamma_{jk}=\begin{cases}1, & \text{第 } j \text{ 个观测来自第 } k \text{ 个分模型}\\ 0, & \text{否则}\end{cases}$$

EM算法在高斯混合模型学习中的应用

- 1、明确隐变量，写出完全数据的对数似然函数：
- 完全数据：　$(y_j, \gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jK})$，　$j = 1, 2, \cdots, N$
- 似然函数：　$P(y, \gamma \mid \theta) = \prod_{j=1}^{N} P(y_j, \gamma_{j1}, \gamma_{j2}, \cdots, \gamma_{jK} \mid \theta)$

$$= \prod_{k=1}^{K} \prod_{j=1}^{N} \left[\alpha_k \phi(y_j \mid \theta_k)\right]^{\gamma_{jk}}$$

$$n_k = \sum_{j=1}^{N} \gamma_{jk} \qquad = \prod_{k=1}^{K} \alpha_k^{n_k} \prod_{j=1}^{N} \left[\phi(y_j \mid \theta_k)\right]^{\gamma_{jk}}$$

$$\sum_{k=1}^{K} n_k = N \qquad = \prod_{k=1}^{K} \alpha_k^{n_k} \prod_{j=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(y_j - \mu_k)^2}{2\sigma_k^2}\right)\right]^{\gamma_{jk}}$$

EM算法在高斯混合模型学习中的应用

- 1、明确隐变量，写出完全数据的对数似然函数：

$$\log P(y,\gamma|\theta)=\sum_{k=1}^{K}n_k\log\alpha_k+\sum_{j=1}^{N}\gamma_{jk}\left[\log\left(\frac{1}{\sqrt{2\pi}}\right)-\log\sigma_k-\frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]$$

EM算法在高斯混合模型学习中的应用

- 2、EM算法的E步，确定Q函数

$$Q(\theta,\theta^{(i)}) = E[\log P(y,\gamma|\theta)|y,\theta^{(i)}]$$
$$= E\left\{\sum_{k=1}^{K} n_k \log\alpha_k + \sum_{j=1}^{N}\gamma_{jk}\left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \log\sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]\right\}$$
$$= \sum_{k=1}^{K}\left\{\sum_{j=1}^{N}(E\gamma_{jk})\log\alpha_k + \sum_{j=1}^{N}(E\gamma_{jk})\left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \log\sigma_k - \frac{1}{2\sigma_k^2}(y_j-\mu_k)^2\right]\right\}$$

需要计算 $E(\gamma_{jk}|y,\theta)$，记为 $\hat{\gamma}_{jk}$

- 第j个观测数据来自第k个分模型的概率，称为分模型k对观测数据y_j的响应度。

EM算法在高斯混合模型学习中的应用

- 2、EM算法的E步，确定Q函数

$$
\begin{aligned}
\hat{\gamma}_{jk} &= E(\gamma_{jk} \mid y, \theta) = P(\gamma_{jk} = 1 \mid y, \theta) \\
&= \frac{P(\gamma_{jk} = 1, y_j \mid \theta)}{\sum_{k=1}^{K} P(\gamma_{jk} = 1, y_j \mid \theta)} \\
&= \frac{P(y_j \mid \gamma_{jk} = 1, \theta) P(\gamma_{jk} = 1 \mid \theta)}{\sum_{k=1}^{K} P(y_j \mid \gamma_{jk} = 1, \theta) P(\gamma_{jk} = 1 \mid \theta)} \\
&= \frac{\alpha_k \phi(y_j \mid \theta_k)}{\sum_{k=1}^{K} \alpha_k \phi(y_j \mid \theta_k)}, \quad j = 1, 2, \cdots, N; \quad k = 1, 2, \cdots, K
\end{aligned}
$$

EM算法在高斯混合模型学习中的应用

- 2、EM算法的E步，确定Q函数

将 $\hat{\gamma}_{jk} = E\gamma_{jk}$ 及 $n_k = \sum_{j=1}^{N} E\gamma_{jk}$ 代入

$$Q(\theta, \theta^{(i)}) = \sum_{k=1}^{K} n_k \log \alpha_k + \sum_{k=1}^{N} \hat{\gamma}_{jk} \left[\log\left(\frac{1}{\sqrt{2\pi}}\right) - \log \sigma_k - \frac{1}{2\sigma_k^2}(y_j - \mu_k)^2 \right]$$

EM算法在高斯混合模型学习中的应用

- 3、确定EM算法的M步：
- 求： $$\theta^{(i+1)} = \arg\max_{\theta} Q(\theta, \theta^{(i)})$$

  用 $\hat{\mu}_k$，$\hat{\sigma}_k^2$ 及 $\hat{\alpha}_k$，$k = 1, 2, \cdots, K$，表示 $\theta^{(i+1)}$

- 采用求导的方法：

$$\hat{\mu}_k = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk} y_j}{\sum_{j=1}^{N} \hat{\gamma}_{jk}} \qquad \hat{\sigma}_k^2 = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk} (y_j - \mu_k)^2}{\sum_{j=1}^{N} \hat{\gamma}_{jk}} \qquad \hat{\alpha}_k = \frac{n_k}{N} = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk}}{N}$$

高斯混合模型参数估计的EM算法

- 输入：观测数据$y_1,y_2,\ldots y_N$, 高斯混合模型
- 输出：高斯混合模型参数
- 1、设定初始值开始迭代
- 2、E步，响应度计算

$$\hat{\gamma}_{jk}=\frac{\alpha_k\phi(y_j\mid\theta_k)}{\sum_{k=1}^{K}\alpha_k\phi(y_j\mid\theta_k)}$$

高斯混合模型参数估计的EM算法

- 输入：观测数据$y_1,y_2,\ldots y_N$, 高斯混合模型
- 输出：高斯混合模型参数
- 3、M步，计算新一轮迭代的模型参数：

$$\hat{\mu}_k = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk} y_j}{\sum_{j=1}^{N} \hat{\gamma}_{jk}} \qquad \hat{\sigma}_k^2 = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk} (y_j - \mu_k)^2}{\sum_{j=1}^{N} \hat{\gamma}_{jk}} \qquad \hat{\alpha}_k = \frac{\sum_{j=1}^{N} \hat{\gamma}_{jk}}{N}$$

- 4、重复2，3步直到收敛

• Q & A

七月在线
JULYEDU.COM
从零学AI的路线图：挑战年薪40万
通过七月在线《机器学习集训营》可解锁全部技能　超过2000人的成功经验：从零学AI、传统IT转型AI
Python基础
核心语法：循环 判断 控制
函数与面向对象
迭代器、生成器、装饰器
数据分析
科学计算之numpy
数据分析之pandas
数据分析实战（美国大选、房价预测）
数据结构
链表、队列、堆栈
字符串和数组
哈希表、树、图
查找与排序：增删改查
分治递归回溯
贪心和动态规划
概率与组合
大数据
Hadoop 基础(HDFS与YARN)
MapReduce与Hive SQL
分布式数据库Hbase
Spark与Flink
机器学习
机器学习的基本流程
经典模型：线性模型、决策树
常考模型：SVM与XGBoost
应用模型：HMM与CRF
特征工程：数据处理、模型构建/调优
上线部署：模型调参与模型评估
项目实战：图像检索、金融风控
深度学习
核心模型：CNN RNN LSTM
CV应用：Two-Stage和One-Stage框架
项目实战：调参、优化、模型压缩、蒸馏收敛
框架应用：TensorFlow与Pytorch
CV NLP 推荐的企业级项目实战
大规模跨境追踪/重识别（ReID）
人体关节点提取
智能客服系统
人体关节点提取
电商平台的商品推荐系统
从零开整电影推荐网站
简历指导、面试辅导、就业内推
CV 企业级项目实战
NLP 企业级项目实战
推荐的企业级项目实战
阶段一 夯实AI基础：数据分析与数据结构
阶段二 掌握AI核心：大数据/ML/DL
阶段三 CV NLP 推荐项目实战