监督学习算法

- EM（期望极大算法）
  - 是一种迭代算法，用于含有隐变量的概率模型的参数的极大似然估计，或极大后验估计。
    - 知道样本求模型参数，用极大似然法或者贝叶斯估计法；假如在这个过程中，还有一个隐变量，模型对它有依赖，此时就要用EM算法了。
  - EM算法步骤
    - E步，求期望
      - 1.给θ自主规定个初值,开始迭代；
      - 2.假设θ_i为第i次迭代参数θ的估计值。根据给定观测数据和当前的参数θ_i，求完全数据的对数似然函数logP(Y,Z|θ)关于隐变量z的条件概率分布P(Z|Y,θ_i)的期望——Q(θ,θ_i)；
        - 完全数据的对数似然函数logP(Y,Z|θ)关于Z的条件概率分布P(Z|Y,θ_i)的期望，称为Q函数。
          - 这个Q函数是由似然函数L(θ)=logP(Y|θ)的下界B省去无关紧要的常数项推导出来的。在迭代的过程中，求得的参数θ的序列使得下界B不断增加，来保证观测数据的对数似然函数在每次迭代中也不断增大。
    - M步，求极大
      - 3.求第二步得到的的函数Q(θ,θ_i)对θ的极大值，作为i+1次迭代的θ估计值；
      - 4.重复第二步和第三步，直到收敛。
  - 注意点
    - EM值对初值很敏感，不同的初值得到的参数估计序列θ_i不一样。
    - EM算法每次迭代后均提高观测数据的似然函数值。
    - 参数估计序列收敛到对数似然函数序列的稳定点，不能保证收敛到极大值点
    - 一般情况下，EM算法是收敛的，但不能保证收敛到全局最优
  - 举例子：两枚硬币A和B，正面朝上的概率分别为PA和PB，抛硬币是如果我们不知道抛的硬币是A还是B，则A和B是隐变量。以这个例子说明EM算法思想。（凑合帮助理解吧）
    - 给PA和PB一个初值
    - 根据这个初值，以及目前观测到的每一轮正反正反这样的结果，算出每一轮是硬币A以及硬币B的概率（求期望 E步），哪个概率大可能就是哪个硬币（最大似然的思想），这就相当于确定了Z值。
    - 现在相当于知道了Z值，这一步根据z值，在结合观测量，算出新的PA和PB（知道结果求参数）M步
    - 反复进行，直到收敛
  - EM算法的应用
    - 估计高斯混合模型的参数
      - 明确隐变量，写出完全数据的对数似然函数
      - EM算法的E步：确定Q函数
      - EM算法的M步：求Q函数对θ的极大值
      - 继续下一轮迭代，直到收敛。
    - 隐马尔可夫模型的无监督学习
- HMM（隐马尔可夫模型）
  - 隐马尔可夫模型模型的三要素
    - A：状态转移概率矩阵
    - B：观测概率矩阵
    - π：初始状态概率向量
  - 该算法解决三个基本问题
    - 已知模型，求解产生某个观测序列的概率（概率计算问题）
      - 直接计算
        - 概念上可行，计算上不可行
      - 前向算法
        - 基于前向概率的定义和递推公式
      - 后向算法
        - 基于后向概率的定义和递推公式
    - 知道观测序列，求模型λ=(A,B,π)参数，使得在该模型下观测序列概率P(O|λ)最大。即用极大似然估计法估计参数。（学习问题）
      - Baul-Welch算法（即EM算法）
        - 无监督学习算法
    - 知道模型和观测序列，求最有可能对应的状态序列，即求给定观测序列条件概率P(I|O)最大的状态序列（预测问题，或叫解码问题）
      - 维比特算法
        - 应用动态规划求解最优路径，即概率最大的状态序列