- 监督学习算法
  - 最大熵模型 (二分类或多分类)
    - 条件熵(模型目标)
      - $$H(Y|X)=H(P_{条})=\sum_{i=1}^{n}p_iH(Y|X=x_i)$$ $$=\sum_x p(x)(-\sum_y p(y|x)logp(y|x))$$ $$=-\sum_{x,y}\tilde{p}(x)p(y|x)logp(y|x)$$
        - 熵只依赖于分布，而与取值无关,，所以可以写为H(P)
        - 熵的范围 :0<=H(X)<=logn
          - 均匀分布时，右边的等号成立
            - （n为X的取值个数）
        - 目标函数中含有log，所以与LR一样都是对数线性模型
    - 条件分布 P(Y|X) 的熵最大的模型
      - 约束条件是：
两个期望相等；
条件概率的和为1
        - $$E_{\tilde{p}}(f_i)=E_p(f_i) \qquad i=1,2,\dots,n$$ 约束条件构成一个集合C，有n个特征函数（代表x,y之间满足的某一种事实），就有n个约束条件
          - $$E_{\tilde{p}}(f)=\sum_{x,y}\tilde{p}(x,y)f(x,y)$$ $$E_p(f)=\sum_{x,y}\tilde{p}(x)p(y|x)f(x,y)$$ 两个期望相等：一个用的是联合分布的经验分布，一个用的是边缘分布的经验分布
        - $$\sum_y p(y|x)=1$$ 条件概率的和为1
    - 算法思想及原理
      - 最大熵模型认为，学习概率模型时，在所有可能的概率模型中，熵最大的模型是最好的模型。首先应当满足所有的约束条件（先验信息），进而对未知的情况不做任何的主观假设，即剩下的部分让他们均匀分布，根据上面的不等式，均匀分布时熵最大，通过熵的最大化来表示等可能性。
      - 模型的学习是对模型进行极大似然估计，或正则化的极大似然估计
        - 模型学习的最优化算法
          - 改进的迭代尺度法（IIS）
          - 牛顿法或拟牛顿法（BFGS）
    - 优点
      - 1.最大熵统计模型获得的是所有满足约束条件的模型中信息熵极大的模型,作为经典的分类模型时准确率较高。
      - 2.可以灵活地设置约束条件，通过约束条件的多少可以调节模型对未知数据的适应度和对已知数据的拟合程度
    - 缺点
      - 由于约束条件的个数往往是跟样本的数量有关，因此当样本数量越来越多的时候，对应的约束条件也会相应增加，这样就会导致计算量越来越大，迭代速度越来越慢，这在实际应用中很难。
    - 计算过程用到了：拉格朗日对偶性
      - 拉个朗日一般求的是最小问题，如果求的是最大就转换一下，加个负号
      - 原始无约束问题就是极小极大问题，关于x的函数，先固定x，求解乘子的解，再确定x
      - 对偶问题是极大极小问题，是关于拉格朗日乘子的函数，先固定乘子，求最优化x的解，再确定乘子。
      - 原始问题和对偶问题最优解相等的条件是KKT条件
        - KKT条件是什么？
          - 原始拉格朗日函数求导为0
          - 原始不等式条件
          - 拉格朗日算子条件以及对偶互补条件
  - 决策树（分类与回归） 31