# 人工智能原理5

2024年6月23日 17:33

# -----机器学习------

#### 1. 监督学习

### (1) 决策树(不都是线性分类器,可以表示非线性函数)

- 随机变量的熵(更大的熵意味着更少的信息,更多的不确定性):  $H(V)=-\sum p(v_k)log_2p(v_k)$
- 布尔随机变量的熵: B(q)与上面的类似
- 属性A的信息增益: Gain(A)=B(p/p+n)-Reminder(A)
- 测试剩余熵: Reminder (A) =  $\sum_{k=0}^{n} \left(\frac{p_k + n_k}{p + n}\right) B\left(\frac{p_k}{p_k + n_k}\right)$
- 决策树将选择有最大信息增益的结点
- 拓展: 随机森林, XGBoost

#### (2) 线性回归与分类

- 回归: 最小化均方误差, 略
- 分类: 感知器学习规则。对于数据不可分离的情况,采用Sigmoid函数进行概率化表示。对于多分类问题,采用Softmax函数表示出每一个类别的概率。最后都是最大似然估计。

$$\max_{w} \ ll(w) = \max_{w} \ \sum_{i} \log P(y^{(i)}|x^{(i)}; w)$$

# 感知器学习规则

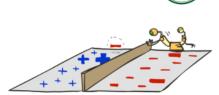


- 如果真实的 y ≠ hw(x) (出错), 调整权重
- 如果 w·x < 0 但输出应是 y=1
  - False negative假阴性
  - x<sub>i</sub>为正时w<sub>i</sub>增加 (为了使输出为1,应使w·x更大)
  - · x<sub>i</sub>为负时w<sub>i</sub>减小
- 如果 w·x > 0 但输出应是 y=0
  - False positive误报
  - x;为正时w;减小(为了使输出为0,应使w·x更小)
  - · x<sub>i</sub>为负时w<sub>i</sub>增加
- 什么样的参数更新规则可以反映上面的思想呢?
- 感知器学习规则:
  - $\mathbf{w} \leftarrow \mathbf{w} + \alpha (\mathbf{y} \mathbf{h}_{\mathbf{w}}(\mathbf{x})) \mathbf{x}$

learning rate 学习速率

+1, -1, or 0 (no error)

• 支持向量机SVM、最近邻等(属于非参数模型)。



# (3) 训练策略

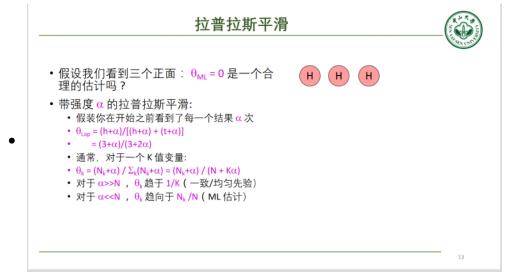
• 随机梯度(SGD), 批梯度(BGD), 小批量梯度(MBGD)、正则化

#### 2. 神经网络

- 全连接、CNN、RNN、ResNet、Attention、Transformer
- 关于梯度下降的算法参考"深度学习"--"梯度下降"

## 3. 统计学习和贝叶斯网络

- 极大似然估计
- 拉普拉斯平滑



• 朴素贝叶斯