

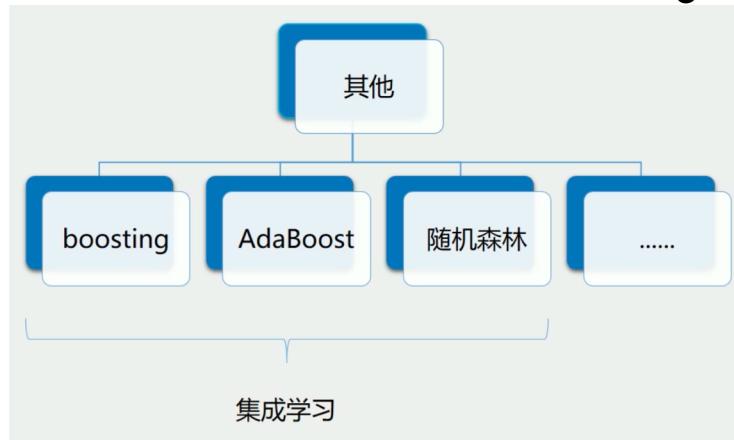
机器学习绪论&数学基础

www



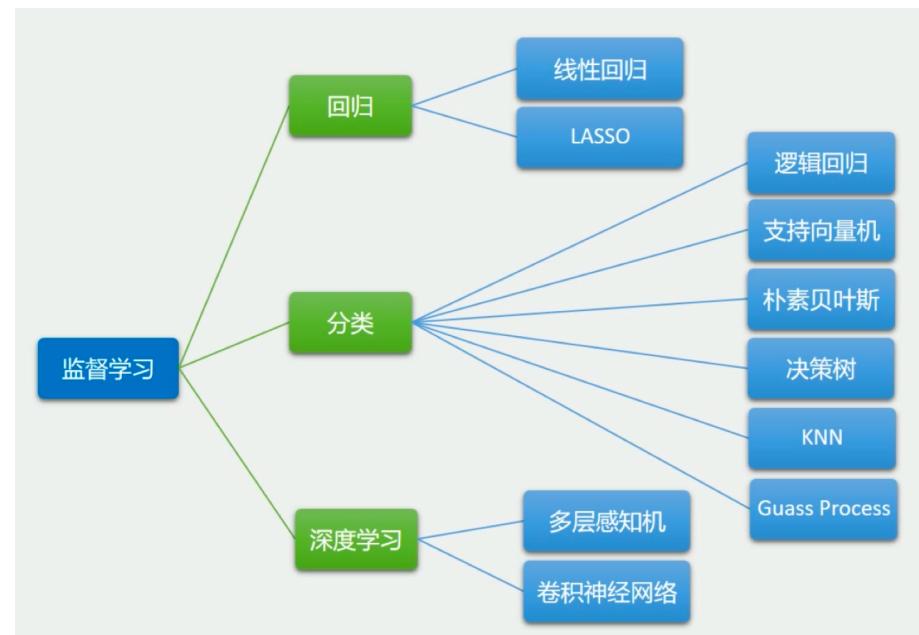
机器学习的分类

- 监督学习 Supervised Learning
- 非监督学习 Unsupervised Learning
- 半监督学习 Semi-Supervised Learning
- 增强学习 Reinforcement Learning



监督学习

- 每组训练数据都有已知的特征(feature)和对应的标签(label)
- 目标：通过训练数据，寻找正确的特征与标记之间的对应关系



REGRESSION / CLASSIFICATION

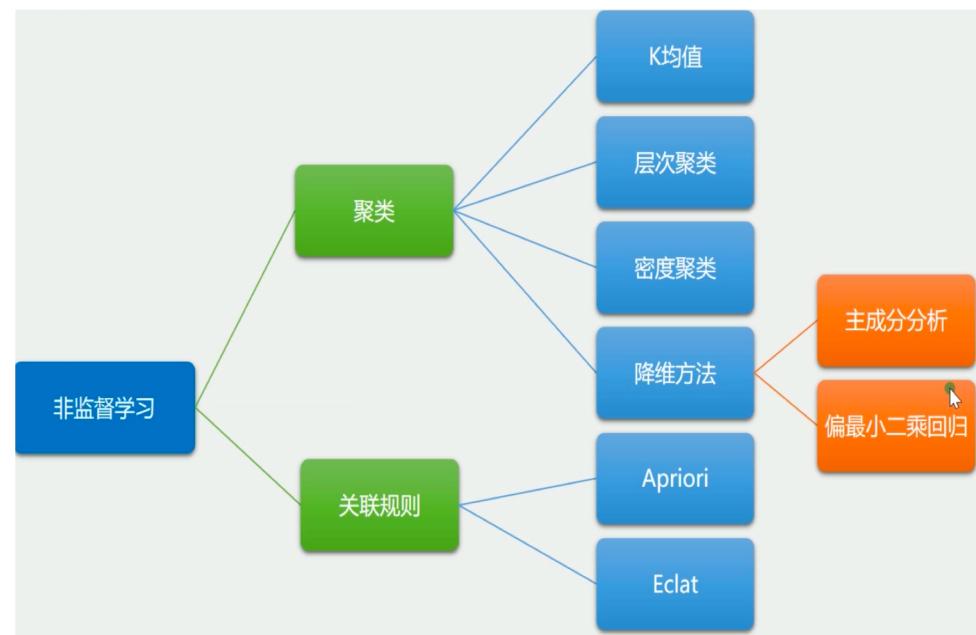
In a **regression** problem, we are trying to predict results within a **continuous** output, meaning that we are trying to map input variables to some **continuous** function.

In a **classification** problem, we are instead trying to predict results in a **discrete** output. In other words, we are trying to map input variables into **discrete** categories.



非监督学习

- 每组训练数据都有已知的特征，但没有任何标记
- 意义：在实际应用中，标签的获取常常需要极大的人工工作量，有时甚至非常困难
- 主要研究聚类问题(clustering)：
 - 按照特征将相似数据聚集在一起



半监督学习

- 训练数据的一部分是有标签的，另一部分没有标签；
- 现实中，没标签数据的数量常常极大于有标签数据数量。
- 隐藏在半监督学习下的基本规律在于：数据的分布必然不是完全随机的，通过一些有标签数据的局部特征，以及更多没标签数据的整体分布，就可以得到可以接受甚至是非常好的分类结果。



增强学习

- **让智能体在环境里学习。** 每个行动会对应各自的奖励(reward)或惩罚(punish)。智能体通过分析数据来学习，怎样的情况下应该做怎样的事情。



其实，这样的学习过程和我们自然的经历非常相似。想象自己是个小孩子，第一次看到了火，然后走到了火边。

你感受到了温暖。火是个好东西 (+1)。



然后就试着去摸。卧槽，这么烫 (-1)。



结论是，在稍远的地方火是好的，靠得太近就不好。



微积分、线代、统计概率



优化基础

- 极大值，极小值，最大值最小值
- 正定矩阵，负定矩阵
- 梯度，黑塞矩阵
- 对一个二元二次求解找最值
- 凹凸性 – 函数凹凸性
- 梯度下降法 目的，步骤（数学原理略）
- 牛顿法 目的，步骤（数学原理略）
- 两种方法的比较



梯度下降法与牛顿法对比

梯度下降法

原理 利用一阶泰勒公式，选取负梯度方向作为下降方向 $\Delta \mathbf{x}^{(k)} = -\nabla f(\mathbf{x}^{(k)})$ ，并计算步长 t_k ，得到 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t_k \Delta \mathbf{x}^{(k)}$

优点 只需要计算梯度，适用于所有一阶可微函数

缺点 收敛速度较慢，迭代次数较多！

牛顿法

利用二阶泰勒公式和极值原理，计算梯度和黑塞矩阵，得到 $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \nabla^2 f(\mathbf{x}^{(k)})^{-1} \nabla f(\mathbf{x}^{(k)})$

利用梯度和黑塞矩阵的信息，迭代次数更少
不需要计算步长

需要计算黑塞矩阵，只适用于二阶可微函数，并且
要求黑塞矩阵可逆。

