

半监督学习

1. 什么是半监督学习？请简要描述其定义和主要思想。它与监督学习和无监督学习有什么区别？

半监督的学习：有少量训练样本，学习机以从训练样本获得的知识为基础，结合测试样本的分布情况逐步修正已有知识，并判断测试样本的类别。

有监督的学习：学习器通过对大量有标记的训练例进行学习，从而建立模型用于预测未见示例的标记，很难获得大量的标记样本。

无监督的学习：无训练样本，仅根据测试样本的在特征空间分布情况进行标记，准确性差。

2. 解释图半监督学习（graph-based semi-supervised learning）的基本思想和过程。

基于相似的样本具有相同标签的平滑假设，构建图结构，利用节点标签扩散，将已标记样本根据样本特征的相似性进行扩散，获得未标记样本的标记。过程包括：图表示、初始化标签、标签传播、学习模型、分类预测。

3. 什么是半监督聚类？它与传统聚类方法有何不同？

半监督聚类在拥有部分额外监督信息时，可利用监督信息改善聚类效果，结合了有标签的数据和未标签的数据。传统聚类方法通常只使用无标签数据进行聚类。

4. 什么是“伪标签”？它在半监督学习中的作用是什么？

"伪标签"是一种半监督学习中的策略，它通过使用已训练模型对未标记数据进行预测，并将这些预测结果作为虚拟的标签（伪标签）来扩充训练数据。可以扩充训练数据和引入额外信息，但是需要注意误差传播的问题。