

机器学习里有哪些分类模型

机器学习任务可以大体上分为两大类：一类是预测型的，基于观察数据对类别或其它变量进行预测，主要包括分类任务和回归任务；另一类是描述型的，通过估计数据的概率分布来描述数据的自有特性，主要包括聚类任务和流形学习任务。

在这些任务中，分类具有重要地位，我们日常生活中遇到的问题大部分是分类问题，如识别车牌号、识别人脸、发布极端天气预警等。

分类模型有很多，总结起来可以分为两种，一种是生成式分类模型，一种是区分性分类模型。

如图 1 (a) 所示，生成式分类模型的基本思路是对每一类数据建立一个生成模型，用以描述该类数据的分布特性；对新来的测试样本（图中红色三角），哪一类数据对应的模型生成该测试样本的概率最大，说明该样本和这一类数据更相似，因此应该被分到这一类中。常见的生成式分类模型包括高斯模型和高斯混合模型等。

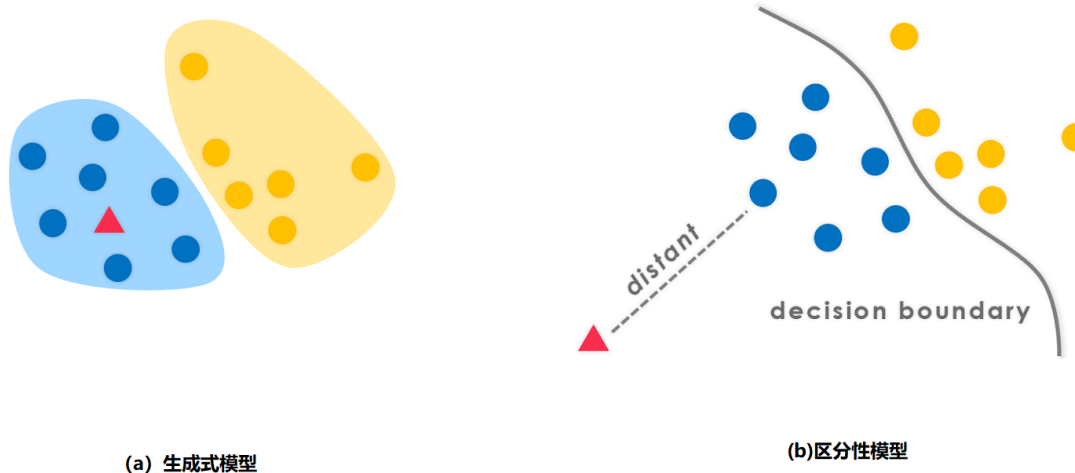


图 1：生成式分类模型和区分性分类模型

区分性分类模型如图 1 (b) 的所示。和生成式模型不同，区分性模型不对每一类数据建模，而是关注分类面，把分类面确定下来（如图中黑色曲线所示），所有数据所属的类别也就确定下来了。常见的区分性分类模型包括支持向量机（SVM）和人工神经网络（ANN）。

总体来说，生成式分类模型具有较强的数据分布假设，当该假设与数据实际分布一致时，可取得很好的性能，同时，模型训练所需数据较少，相应的缺点是学习能力较弱，处理复杂数据能力有限；反之，区分性模型对数据分布假设少，学习能力强，但需要大量训练数据。当训练数据量足够多时，灵活的区分性模型可取得非常好的效果，这也是深度神经网络近年来取得长足进步的原因之一。