### 3.2.2 常见的模型分类方法

比较通用的模型分类方法是根据训练样本是否带有标签分为监督学习和非监督学习。

**Q**：阐述监督学习和非监督学习的区别。

训练数据既有特征（feature），又有标签（label），则称为监督学习。通过训练，让机器可以自己找到特征和标签之间的联系，针对只有特征没有标签的数据，即此前提到的测试集，可以通过模型获得标签。

根据标签是连续的或者离散的，分为预测（prediction）问题和分类（classification）问题。需要注意的是，这里的离散和连续的区分依据是标签数量是否可数，而非是否有限（关于可数，在前面的章节中介绍过）。

在非监督学习的数据集中只有特征，没有标签，通过数据之间的内在联系和相似性将样本划分成若干类，称为聚类（clustering），或者对高维数据进行降维（dimension reduction）。

这里需要搞清楚分类和聚类的区别。分类是指在监督学习中，在标签可数的情况下判断结果所属的类别；而聚类则是指在非监督学习中，通过数据之间的内在联系和相似性将样本划分成若干类。

根据上述分类方法，对一些常见的模型进行分类，如表 3-5 所示。

表 3-5

| 监督学习 | 非监督学习 |
| --- | --- |
| 预测问题：线性回归模型、时间序列模型、神经网络模型 | 聚类问题：K-Means 聚类模型、DBSCAN 聚类模型、E-M 聚类模型 |
| 分类问题：逻辑回归模型、SVM 模型、决策树模型、随机森林模型、Boosting 模型 | 降维问题：PCA（主成分分析法）模型 |

在数据挖掘中，模型也可以分为参数模型和非参数模型。