

3.判别式模型和生成式模型

对于有监督学习可以将其分为两类模型：判别式模型和生成式模型。简单地说，判别式模型是针对条件分布建模，而生成式模型则针对联合分布进行建模。

1.基本概念

假设我们有训练数据(X,Y)，X是属性集合，Y是类别标记。这时来了一个新的样本x，我们想要预测它的类别y。

我们最终的目的是求得最大的条件概率 $P(y|x)$ **【在特征是x的条件下标签是y的概率】**作为新样本的分类。

1.1 判别式模型这么做：

根据训练数据得到分类函数和分界面，比如说根据SVM模型得到一个分界面，然后直接计算条件概率 $P(y|x)$ ，我们将最大的 $P(y|x)$ 作为新样本的分类。

判别式模型是对条件概率建模，学习不同类别之间的最优边界，无法反映训练数据本身的特性，能力有限，其只能告诉我们分类的类别。

1.2 生成式模型这么做

一般会对每一个类建立一个模型，有多少个类别，就建立多少个模型。比如说类别标签有{猫，狗，猪}，那首先根据猫的特征学习出一个猫的模型，再根据狗的特征学习出狗的模型，之后分别计算新样本x跟三个类别的联合概率 $P(x,y)$ ，然后根据贝叶斯公式：

$$P(y|x) = \frac{P(x,y)}{P(x)}$$

分别计算 $P(y|x)$ ，选择三类中最大的 $P(y|x)$ 作为样本的分类。

1.3 两个模型的小结

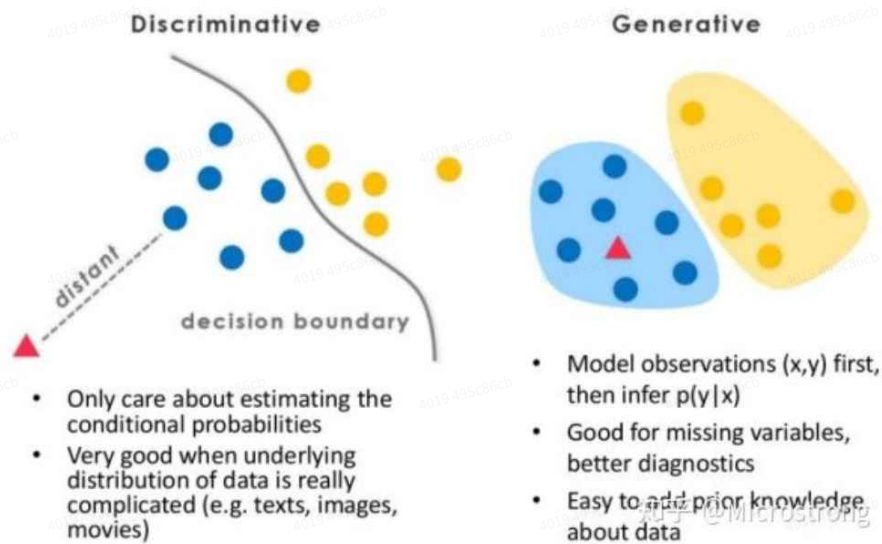
不管是生成式模型还是判别式模型，它们最终的判断依据都是条件概率 $P(y|x)$ 。

但是生成式模型先计算了联合概率 $P(x,y)$ ，再由贝叶斯公式计算得到条件概率。因此，生成式模型可以体现更多数据本身的分布信息，其普适性更广。

2.两者区别

2.1 判别式模型和生成式模型的对比图

Discriminative vs. Generative



上图左边为判别式模型而右边为生成式模型，可以很清晰地看到差别，**判别式模型是在寻找一个决策边界，通过该边界来将样本划分到对应类别。**

而生成式模型则不同，**它学习了每个类别的边界，它包含了更多信息，可以用来生成样本。**

2.2两者所包含的算法

机器学习

判别式模型

- 线性回归 (Linear Regression)
- 逻辑回归 (Logistic Regression)
- 线性判别分析
- 支持向量机 (SVM)
- CART (Classification and Regression Tree)
- 神经网络 (NN)
- 高斯过程 (Gaussian Process)
- 条件随机场 (CRF)

生成式模型

- 朴素贝叶斯
- K近邻 (KNN)
- 混合高斯模型
- 隐马尔科夫模型 (HMM)
- 贝叶斯网络
- Sigmoid Belief Networks
- 马尔科夫随机场 (Markov Random Fields)
- 深度信念网络 (DBN)
- LDA文档主题生成模型

@Microstrong