# 大作业05-关于概率图模型的学习

Matthew

2023年05月29日

**1 分别以朴素贝叶斯和条件随机场为例，回答在概率图模型的训练中，什么是生成式训练、判别式训练，及其优缺点？（10分）。注意：在在本课程中不需要考虑模型是决策函数，而只需要考虑概率模型。**

**生成式训练：**

生成式训练是基于生成模型的一种训练方法。生成模型试图学习样本的联合概率分布，并通过使用贝叶斯规则来计算给定输入的后验概率。在朴素贝叶斯中，生成式训练通过学习类别的先验概率和特征的条件概率来构建模型。它假设输入特征与类别之间的条件概率独立，并利用训练数据中的样本频率进行参数估计。生成式模型可以用于生成新的样本数据。

生成式训练的优点：

1 数据建模能力强：生成模型可以对数据的分布进行建模，能够表示样本的整体分布特征。

2 可以处理缺失数据：生成模型能够自然地处理缺失数据，因为它们对整个数据分布进行建模。

生成式训练的缺点：

1 对数据分布的假设限制较多：生成模型通常会对输入特征的分布进行假设，这可能会导致对数据的一些偏见。

2 在分类任务中，生成式模型可能无法对类别边界进行建模，因为它们更关注数据分布的整体特征。

**判别式训练：**

判别式训练是基于判别模型的一种训练方法。判别模型直接对条件概率分布进行建模，学习输入特征与输出标签之间的映射关系。在条件随机场中，判别式训练通过学习特征函数和标签序列之间的条件概率来构建模型。它不对输入特征的分布进行假设，而是关注于最大化给定输入下输出标签的条件概率。

判别式训练的优点：

1 对分类任务更有效：判别模型直接对条件概率进行建模，可以更好地拟合类别之间的边界，对于分类任务的性能表现通常更好。

2 可以灵活地引入特征：判别模型可以根据任务的需要选择特征，可以引入更多的上下文信息来提高模型性能。

判别式训练的缺点：

1 对缺失数据处理困难：判别模型对于缺失数据的处理相对困难，因为它们需要已知输入特征来预测输出标签。

2 可能对训练数据敏感：判别模型对训练数据中的噪声和标签错误比较敏感，可能需要更多的数据清洗和预处理工作。

**2 根据下面的训练数据（表1），估计贝叶斯网络的参数。该网络包括随机变量和其均为二值变量，取值为0、1。贝叶斯网络的结构已经给定，如图1所示。采用极大似然估计的方法，估计随机变量和的条件概率分布表，可用分数表达。包括6个参数：**

X=1的频数：11

X=0的频数：9

P(X=1) = 11 / 20 = 0.55

P(X=0) = 9 / 20 = 0.45

X=1, Y=1的频数：6

X=1, Y=0的频数：5

P(Y=1 | X=1) = 6 / 11 ≈ 0.545

P(Y=0 | X=1) = 5 / 11 ≈ 0.455

X=0, Y=1的频数：3

X=0, Y=0的频数：6

P(Y=1 | X=0) = 3 / 9 ≈ 0.333

P(Y=0 | X=0) = 6 / 9 ≈ 0.667