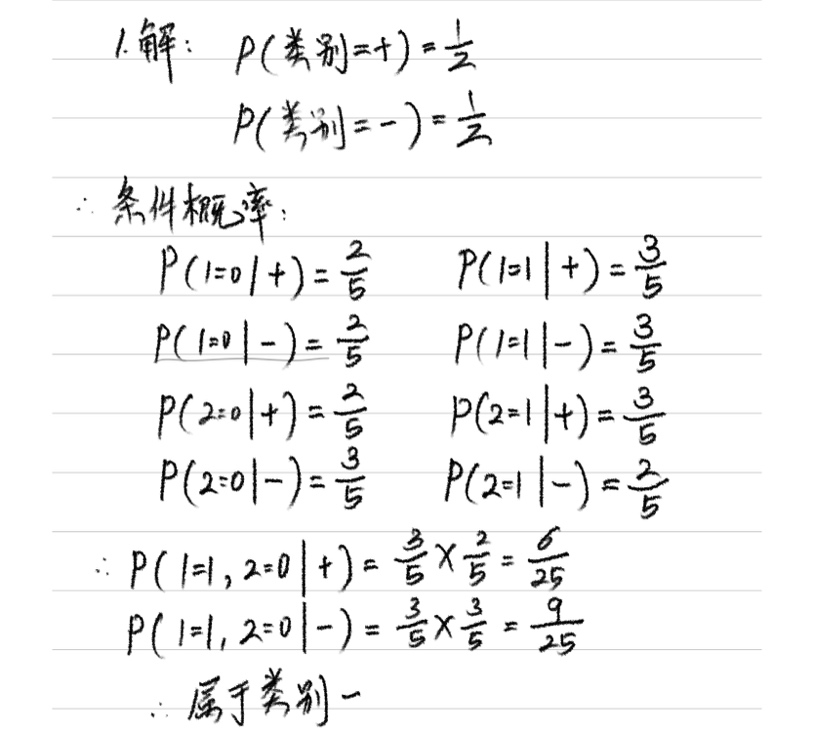
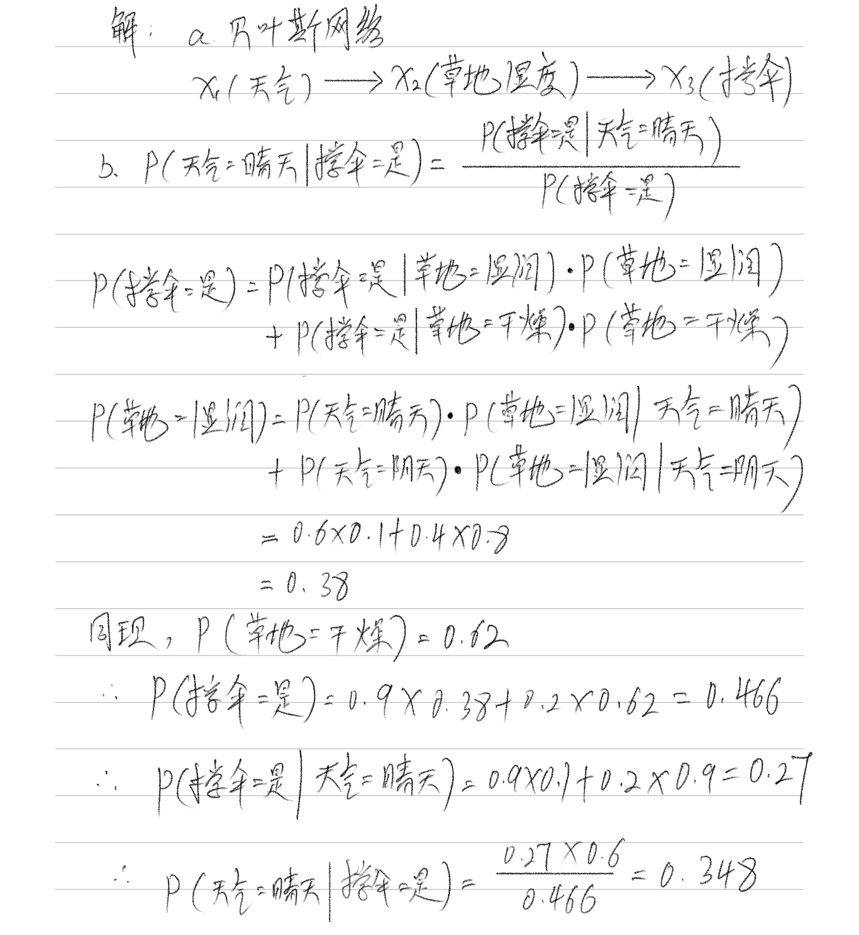
机器学习 作业五

20201060287 李昂

**1. 如下**

****

**2. 如下**

****

**3. 请简述聚类分析和分类有何不同？ 请从算法原理，评价指标等方面展开回答。**

聚类：

1. 目标：聚类分析旨在将数据集中的样本划分为不同的群集或类别，使得同一群集内的样本彼此相似，而不同群集之间的样本差异较大。

2. 算法原理：聚类分析算法基于数据样本之间的相似性或距离来组织数据。常用的聚类算法包括K均值聚类、层次聚类、密度聚类等。这些算法尝试在样本间构建一种相似度度量，然后根据相似度度量将样本划分为不同的群集。

3. 评价指标：常用的聚类评价指标包括轮廓系数、Calinski-Harabasz指数和Davies-Bouldin指数等。这些指标可以衡量聚类结果的紧密度、分离度和聚类质量。

分类：

1. 目标：分类是将数据样本分配到预定义的类别或标签中的过程，即给定一个新样本，将其分配到已知的类别中。

2. 算法原理：分类算法基于训练集中已知的类别信息，通过学习样本的特征和类别之间的关系来构建一个分类模型。常用的分类算法包括决策树、支持向量机、逻辑回归等。这些算法通过在特征空间中定义一个决策边界来将不同类别的样本分开。

3. 评价指标：分类问题的评价指标包括准确率、精确率、召回率、F1分数等。这些指标可以衡量分类模型的性能和预测结果的准确性。

聚类分析和分类在目标、算法原理和评价指标等方面存在明显的不同，适用于不同的数据分析任务。聚类分析用于无监督学习中的数据探索和模式发现，而分类则用于有监督学习中的预测和决策任务。

**4. 标准PCA通过对数据的线性变换来找到主成分，请问如何扩展PCA使其实现非线性降维压缩？**

答：

想要实现非线性降维压缩，可以使用核主成分分析和自编码器。

1. 核主成分分析：

核主成分分析是一种基于核技巧的非线性降维方法，它通过将数据映射到高维空间，然后在高维空间中进行线性降维。它

选择一个适当的核函数（如径向基函数、多项式核函数等）来将原始数据映射到高维特征空间，在高维空间中进行主成分分析，计算数据的主成分，根据主成分的重要性对数据进行降维。核主成分分析允许在非线性数据上进行降维，因为它通过核技巧将数据映射到高维空间，从而可以在高维空间中找到非线性结构的主成分。

2. 自编码器：

自编码器是一种神经网络模型，可以用于非线性降维。它由编码器和解码器两部分组成，其中编码器将输入数据压缩到低维表示，而解码器则将低维表示重构为原始数据。自编码器的训练过程旨在最小化重构误差，使得解码器能够准确地重构原始数据。自编码器可以学习数据中的非线性特征，并将其压缩到低维表示。通过使用编码器部分获取低维表示，可以实现非线性降维。

**5. 请解释贝叶斯网络和马尔可夫随机场之问的区别。**

答：

1. 图结构不同：贝叶斯网络使用有向图，而马尔可夫随机场使用无向图。

2. 条件独立性假设不同：贝叶斯网络假设节点在给定其父节点的情况下与非后代节点条件独立；马尔可夫随机场假设给定与一个节点相邻的所有其他节点的情况下，该节点与其他节点条件独立。

3. 推断问题不同：贝叶斯网络主要用于计算未观测到的节点的后验概率分布；马尔可夫随机场主要用于计算未观测到的变量的条件分布。

**5.请简述强化学习中策略梯度方法与价值函数方法的区别。**

答：

1. 表示方式：策略梯度方法直接对策略进行参数化表示，而价值函数方法通过对状态或状态-动作对的价值进行估计来表示策略的优劣。

2. 目标函数：策略梯度方法的目标是直接优化策略的性能，而价值函数方法的目标是优化价值函数的估计。

3. 更新策略：策略梯度方法通过梯度上升法直接更新策略参数，而价值函数方法通过估计和更新价值函数来间接改进策略。