学习第5章决策树

2019-05-30 00:00 至 2019-05-31 23:59

**Week2**

**Day4-Day5**

**任务简介：**掌握决策树模型的思想和算法。

**详细说明：**第5章介绍的决策树模型即可以解决分类问题也可以解决回归问题，本书重点介绍分类问题。通过学习第1节，了解决策树模型的基本思想；通过学习第2节，了解选择分类特征的两个准则：信息增益和信息增益比；通过学习第3节，理解两个准则下对应的两种决策树算法；通过学习第4节，了解决策树模型中如何通过剪枝控制模型复杂度；通过学习第5节，掌握基尼系数的计算方法和CART算法，其中回归问题不做学习要求。本章的难点是如何理解信息增益和基尼系数以及它们的计算。

**学习目标：**

0.导读视频。

[[](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDQyYTJkYTlmXzdJVVhtd3lXIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDQyYTJkYTlmXzdJVVhtd3lXIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)

[《统计学习方法》决策树](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDQyYTJkYTlmXzdJVVhtd3lXIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)

[[](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDQyYTJkYTlmXzdJVVhtd3lXIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDQyYTJkYTlmXzdJVVhtd3lXIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)

1.理解信息增益、基尼系数的含义和计算方法。

[[](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDU5YmNkMDc5X0FGaU1GVE1BIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDU5YmNkMDc5X0FGaU1GVE1BIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)

[《统计学习方法》信息增益与基尼指数](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDU5YmNkMDc5X0FGaU1GVE1BIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)

[[](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDU5YmNkMDc5X0FGaU1GVE1BIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)](https://appuaaoe86p4947.h5.xiaoeknow.com/content_page/eyJ0eXBlIjoiMiIsInJlc291cmNlX3R5cGUiOjMsInJlc291cmNlX2lkIjoidl81Y2QyZDU5YmNkMDc5X0FGaU1GVE1BIiwiYXBwX2lkIjoiYXBwdWFBb2U4NnA0OTQ3IiwicHJvZHVjdF9pZCI6InBfNWNkZDMyM2UwZWNhZl94MVdoZlJEYiJ9)

2.掌握ID3算法。

3.了解决策树剪枝的目的和剪枝准则。

4. 掌握CART生成算法。

**作业5：**

1.证明CART剪枝算法中，当α确定时，存在唯一的最小子树Ta使损失函数Cα(T)最小。

2.尝试调用sklearn.tree.DecisionTreeClassifier模块，训练数据集采用课本例题5.1的数据，判断是否应该批准下列人员的贷款申请，打卡代码运行结果的截图。

A={青年，否，是，一般}

B={中年，是，否，好}

C={老年，否，是，一般}

作业答案及代码讲解在本周日公布，助教会进行视频讲解。