CIS600 Fundamentals of Data and Knowledge Mining

Week 3 Decision Tree Induction

Discussion Questions

1. What is the supervised learning framework and what are the similarities and dissimilarities among different type of datasets?
2. What are some popular dataset purity measures? How to calculate them?
3. Which criteria does decision tree algorithm use to decide which attribute to split in order to induce a tree?
4. What is the model overfitting phenomenon and how to avoid or reduce overfitting in model building?
5. Hyper-parameters in decision tree algorithm
6. Exercise of growing a decision tree using a toy data set