1. Python

2. Python机器学习的库：scikit-learn

   2.1：特性：
      简单高效的数据挖掘和机器学习分析
      对所有用户开放，根据不同需求高度可重用性
      基于Numpy, SciPy和matplotlib
      开源，商用级别：获得 BSD许可

   2.2 覆盖问题领域：
      分类（classification), 回归（regression), 聚类（clustering), 降维(dimensionality reduction)
      模型选择(model selection), 预处理(preprocessing)

3. 使用用scikit-learn
   安装scikit-learn: pip, easy_install, windows installer
   安装必要package：numpy， SciPy和matplotlib， 可使用Anaconda (包含numpy, scipy等科学计算常用
package）
   安装注意问题：Python解释器版本（2.7 or 3.4？）, 32-bit or 64-bit系统

4. 例子：

RID	age	income	student	credit_rating	Class: buys_computer
1	youth	high	no	fair	no
2	youth	high	no	excellent	no
3	middle_aged	high	no	fair	yes
4	senior	medium	no	fair	yes
5	senior	low	yes	fair	yes
6	senior	low	yes	excellent	no
7	middle_aged	low	yes	excellent	yes
8	youth	medium	no	fair	no
9	youth	low	yes	fair	yes
10	senior	medium	yes	fair	yes
11	youth	medium	yes	excellent	yes
12	middle_aged	medium	no	excellent	yes
13	middle_aged	high	yes	fair	yes
14	senior	medium	no	excellent	no

文档： http://scikit-learn.org/stable/modules/tree.html

解释Python代码

安装 Graphviz： http://www.graphviz.org/
配置环境变量

转化dot文件至pdf可视化决策树：dot -Tpdf iris.dot -o outpu.pdf