属性及取值：  
  
age: continuous.   
workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.   
fnlwgt: continuous.   
education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.   
education-num: continuous.   
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.   
occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.   
relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.   
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.   
sex: Female, Male.   
capital-gain: continuous.   
capital-loss: continuous.   
hours-per-week: continuous.   
native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Dominican-Republic, Laos, Ecuador, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinadad&Tobago, Peru, Hong, Holand-Netherlands.

收入（分类依据）

>50K, <=50K.   
实验说明：

1 对adult.data中的原始数据进行必要的特征选择和预处理:

1） 选择age，marital-status，occupation，capital-gain ，hours-per-week，native-country作为分类属性

2）把native-country归约到大洲名

3）把age归约到几个区间

4）清除原数据中不干净的数据，包括去除噪音数据，填充缺省值

2 以adult.data中选取的数据为样本，以收入为分类依据，建立决策树分类模型，

对训练和测试样本按以下比例分别进行构建分类器：分类器1：(4:1)，分类器2：3:1, 分类器3：9:1

3 画出2中各决策树

4 从adult.test中

1）选取10个样本数据进行利用分类器1进行测试，

2）对数据做必要的预处理和特征选择

3）显示预测结果