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 对原始数据进行必要的预处理，提交报告时，

2 以adult.data中数据为样本，以收入为分类依据，建立分类模型

3 求解分类规则

4 以adult.test为测试数据，对分类器进行评估，并求解以下数值

1. **准确率**
2. 精度
3. 召回率
4. F 度量

设计要求：

1. 实验说明中1的预处理，写明处理过程或方法，及部分原始数据处理后的结果（数据集的5%）
2. 分类模型建立的算法伪代码和源代码（并标注主要过程或方法）
3. 展示20个数据对象的算法执行过程
4. 写出形成的分类规则的源代码（标注主要过程或方法），以及该分类模型形成的分类规则
5. 写出实验说明4）中对前100项测试数据进行测试的结果
6. 基于5）的结果，写出构建的混淆矩阵，并进行实验说明4）的评估，写出指标的求解过程
7. 写出实验说明4）所有测试数据集的指标值