开放工程2: 学生表现预测（10'）

- 分类问题描述
  - (1) 选用33特征（包含类标）、649样本的数据集
  - (2) 将成绩按10%-20%-40%-20%-10%划分为5类
  - (3) 利用前30特征预测G1、G2、G3
  - (4) 利用前32特征预测G3
- 评分指标
  - ☑ 模型必须包含近邻法（2'）、逻辑回归（1vs多or1vs1）（4'）、决策树（2'）
  - ☑ 选用的模型精度、结果对比列表
  - ☑ 讲述清晰度和熟练度（2'）
  - ☑ 上述指标缺少即为不合格

开放工程2: 学生表现预测（10'）

- 数据集： https://archive.ics.uci.edu/ml/datasets/student+performance#

Student Performance Data Set
Download: Data Folder, Data Set Description

Abstract: Predict student performance in secondary education (high school).

Data Set Characteristics:	Multivariate	Number of Instances:	649	Area:	Social
Attribute Characteristics:	Integer	Number of Attributes:	33	Date Donated	2014-11-27
Associated Tasks:	Classification, Regression	Missing Values?	N/A	Number of Web Hits:	1158640

Data Set Information:

This data approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por). In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful (see paper source for more details).

Attribute Information:

Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:
1 school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2 sex - student's sex (binary: 'F' - female or 'M' - male)
3 age - student's age (numeric: from 15 to 22)
4 address - student's home address type (binary: 'U' - urban or 'R' - rural)
5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€" 5th to 9th grade, 3 â€" secondary education or 4 â€" higher education)
8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 â€" 5th to 9th grade, 3 â€" secondary education or 4 â€" higher education)
9 Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
10 Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
11 reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
15 failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16 schoolsup - extra educational support (binary: yes or no)
17 famsup - family educational support (binary: yes or no)
18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19 activities - extra-curricular activities (binary: yes or no)
20 nursery - attended nursery school (binary: yes or no)
21 higher - wants to take higher education (binary: yes or no)
22 internet - Internet access at home (binary: yes or no)
23 romantic - with a romantic relationship (binary: yes or no)
24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29 health - current health status (numeric: from 1 - very bad to 5 - very good)
30 absences - number of school absences (numeric: from 0 to 93)

these grades are related with the course subject, Math or Portuguese:
31 G1 - first period grade (numeric: from 0 to 20)
31 G2 - second period grade (numeric: from 0 to 20)
32 G3 - final grade (numeric: from 0 to 20, output target)

开放工程2: 学生表现预测（10′）

- **数据处理**
  - **标称属性：编码1，2，3...**

student-por

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic	famrel	freetime	goout	Dalc	Walc	health	absences	G1	G2	G3
GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no	no	yes	yes	no	no	4	3	4	1	1	3	4	0	11	11
GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no	no	no	yes	yes	no	5	3	3	1	1	3	2	9	11	11
GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	0	yes	no	no	no	yes	yes	yes	no	4	3	2	2	3	3	6	12	13	12
GP	F	15	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes	no	yes	yes	yes	yes	yes	3	2	2	1	1	5	0	14	14	14
GP	F	16	U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes	no	no	yes	yes	no	no	4	3	2	1	2	5	0	11	13	13
GP	M	16	U	LE3	T	4	3	services	other	reputation	mother	1	2	0	no	yes	no	yes	yes	yes	yes	no	5	4	2	1	2	5	6	12	12	13
GP	M	16	U	LE3	T	2	2	other	other	home	mother	1	2	0	no	no	no	no	yes	yes	yes	no	4	4	4	1	1	3	0	13	12	13
GP	F	17	U	GT3	A	4	4	other	teacher	home	mother	2	2	0	yes	yes	no	no	yes	yes	no	no	4	1	4	1	1	1	2	10	13	13
GP	M	15	U	LE3	A	3	2	services	other	home	mother	1	2	0	no	yes	no	no	yes	yes	yes	no	4	2	2	1	1	1	0	15	16	17
GP	M	15	U	GT3	T	3	4	other	other	home	mother	1	2	0	no	yes	no	yes	yes	yes	yes	no	5	5	1	1	1	5	0	12	12	13
GP	F	15	U	GT3	T	4	4	teacher	health	reputation	mother	1	2	0	no	yes	no	no	yes	yes	yes	no	3	3	3	1	2	2	2	14	14	14
GP	F	15	U	GT3	T	2	1	services	other	reputation	father	3	3	0	no	yes	no	yes	yes	yes	yes	no	5	2	2	1	1	4	0	10	12	13
GP	M	15	U	LE3	T	4	4	health	services	course	father	1	1	0	no	yes	no	yes	yes	yes	yes	no	4	3	3	1	3	5	0	12	13	12
GP	M	15	U	GT3	T	4	3	teacher	other	course	mother	2	2	0	no	yes	no	no	yes	yes	yes	no	5	4	3	1	2	3	0	12	12	13
GP	M	15	U	GT3	A	2	2	other	other	home	other	1	3	0	no	yes	no	no	yes	yes	yes	yes	4	5	2	1	1	3	0	14	14	15
GP	F	16	U	GT3	T	4	4	health	other	home	mother	1	1	0	no	yes	no	no	yes	yes	yes	no	4	4	4	1	2	2	6	17	17	17
GP	F	16	U	GT3	T	4	4	services	services	reputation	mother	1	3	0	no	yes	no	no	yes	yes	yes	no	3	2	3	1	2	2	10	13	13	14
GP	F	16	U	GT3	T	3	3	other	other	reputation	mother	3	2	0	yes	yes	no	no	yes	yes	no	no	5	3	2	1	1	4	2	13	14	14
GP	M	17	U	GT3	T	3	2	services	services	course	mother	1	1	3	no	yes	yes	yes	yes	no	yes	no	5	5	5	2	4	5	2	8	8	7
GP	M	16	U	LE3	T	4	3	health	other	home	father	1	2	0	no	no	no	no	yes	yes	no	no	3	1	3	1	3	5	6	12	12	12
GP	M	15	U	GT3	T	4	3	teacher	other	reputation	mother	1	2	0	no	no	no	no	yes	yes	yes	no	4	4	1	1	1	1	0	12	13	14
GP	M	15	U	GT3	T	4	4	health	health	other	father	1	1	0	no	yes	yes	no	yes	yes	no	no	5	4	2	1	1	5	0	11	12	12
GP	M	16	U	LE3	T	4	2	teacher	other	course	mother	1	2	0	no	no	no	no	yes	yes	yes	no	4	5	1	1	3	5	0	12	13	14
GP	M	16	U	LE3	T	2	2	other	other	reputation	mother	2	2	0	no	no	no	no	yes	yes	yes	no	5	4	4	2	4	5	2	10	10	10
GP	F	15	R	GT3	T	2	4	services	health	course	mother	1	3	0	yes	yes	no	no	yes	yes	yes	yes	4	3	2	1	1	5	2	10	11	10
GP	F	16	U	GT3	T	2	2	services	services	home	mother	1	1	0	no	no	no	no	no	yes	no	no	4	2	2	1	3	5	6	10	11	11
GP	M	15	U	GT3	T	2	2	other	other	home	mother	1	2	0	no	no	no	yes	yes	yes	yes	no	4	2	2	1	3	5	8	11	12	12
GP	M	15	U	GT3	T	4	2	health	services	other	mother				no		no			yes	yes		no	3	2	4	2		5	0	11	11	11

开放工程2: 学生表现预测（10'）

- 有问题请咨询课程教学人员
  - ☑ 有问题请及时联系
- 提交方式
  - ☑ 代码（python（推荐）、或者其他语言）
  - ☑ 文档（300字-500字之间，预测精度对比表格）
  - ☑ 讲解音频（不超过3分钟）

  - ☑ 文件名（同邮件名）：工程2+学号+姓名
  - ☑ 课程邮箱：prml2022@yeah.net
  - ☑ 截止时间：6月5日（儿童节后第一个周日），24:00