CS412 Machine Learning - Homework 1
Decision Tree and K-NN

Deadline: Sunday March 8, 2020, 23:55
Late submission: till Tuesday March 10, 2020, 23:55
(-10pts penalty for **each** late submission day)

Dataset Information

German Credit data from UCI Machine Learning Repository:

`https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)`

The dataset contains 1000 entries with 20 categorical attributes prepared by Prof. Hofmann. Each entry represents a person who takes a credit by a bank. Refer to the dataset website for the attributes details.

Task: Classify a person as good or bad credit risks according to set of attributes.

	A	B	C	D	E	F	G	H	I	J	K	L	
1	id	duration	credit_amount	installment_commitr	residence_since	age	existing_credits	num_dependents	label	f_worker	checking_status_<0	checking_status_>=2	checking_
2	1	6	1169	4	4	67	2	1	1	1	1	0	
3	2	48	5951	2	2	22	1	1	0	1	0	0	
4	3	12	2096	2	3	49	1	2	1	1	0	0	
5	4	42	7882	2	4	45	1	2	1	1	1	0	
6	5	24	4870	3	4	53	2	2	0	1	1	0	
7	6	36	9055	2	4	35	1	2	1	1	0	0	
8	7	24	2835	3	4	53	1	1	1	1	0	0	
9	8	36	6948	2	2	35	1	1	1	1	0	0	
10	9	12	3059	2	4	61	1	1	1	1	0	0	
11	10	30	5234	4	2	28	2	1	0	1	0	0	
12	11	12	1295	3	1	25	1	1	0	1	0	0	
13	12	48	4308	3	4	24	1	1	0	1	1	0	
14	13	12	1567	1	1	22	1	1	1	1	0	0	
15	14	24	1199	4	4	60	2	1	0	1	1	0	
16	15	15	1403	2	4	28	1	1	1	1	1	0	
17	16	24	1282	4	2	32	1	1	0	1	1	0	
18	17	24	2424	4	4	53	2	1	1	1	0	0	
19	18	30	8072	2	3	25	3	1	1	1	1	0	
20	19	24	12579	4	2	44	1	1	0	1	0	0	
21	20	24	3430	3	2	31	1	2	1	1	0	0	
22	21	9	2134	4	4	48	3	1	1	1	0	0	
23	22	6	2647	2	3	44	1	2	1	1	1	0	

Figure 1: Samples from German Credit data.

Startup Code:

`https://drive.google.com/open?id=1e550az93U3_kfRBbVY5PZnMKYwGYmHqi`

To start working for your homework, take a copy of this folder to your own google drive.

Submission

As training and testing may take long time, we will just look at your notebook results; so make sure to run all of the cells and the output results are there.

Please submit your homework as follows:

- Download the .ipynb and the .py file and upload both of them to sucourse.