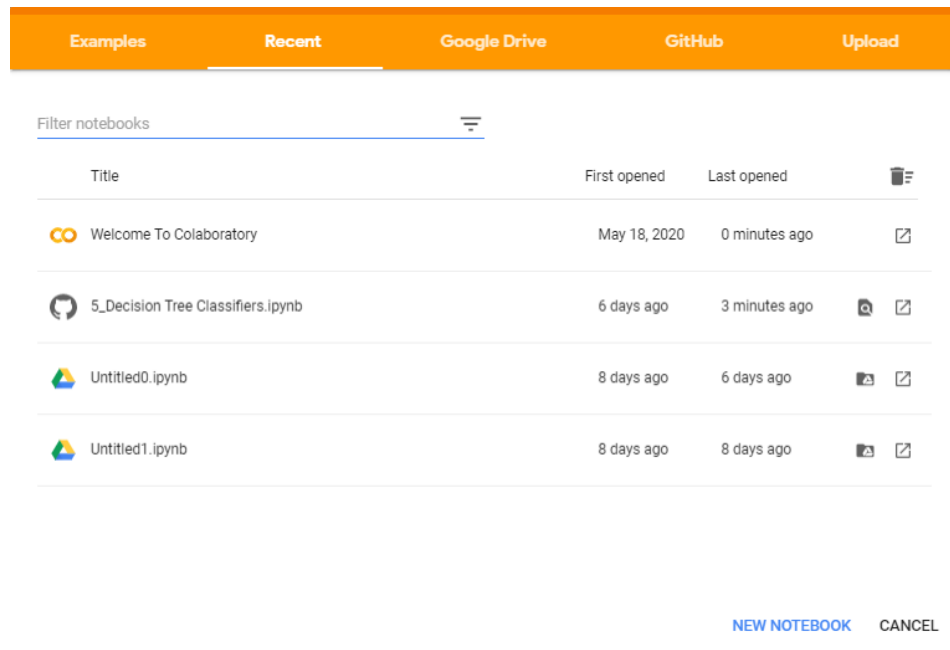
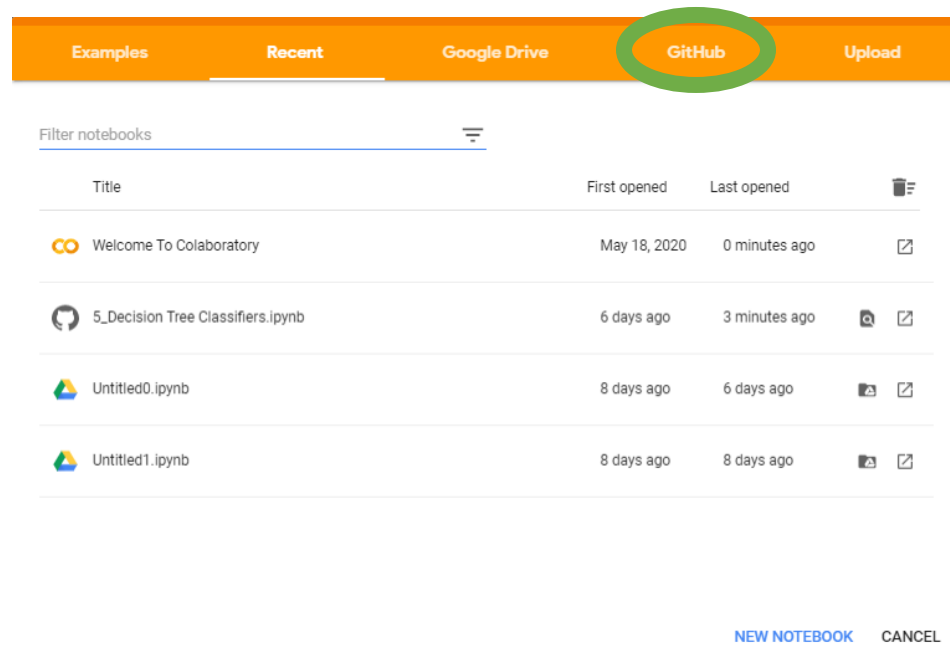


## Getting Started with Google Colab

1. To use Colab you should be logged in to your Google account. After this, click on the following link <https://colab.research.google.com>
2. You will see the following tab pop up when you open Colab. If it doesn't pop up, go to File -> Open Notebook



3. You will be using Notebooks stored on a Github repository. Click on the Github link.



4. Paste the following URL on the search bar:

<https://github.com/khider/INF549> and click enter.

This link should allow you access to all the notebooks on the repository.

Scroll down and select the notebook you need to complete your homework. You should be redirected to Jupyter Notebook.

Examples

Recent

Google Drive

GitHub

Upload

Enter a GitHub URL or search by organization or user

☐ Include private repos

<https://github.com/khider/INF549>

Repository: [khider/INF549](#) Branch: [master](#)

Path

Assignment2\_CaesarCypher\_SimpleMath&Statistics/.ipynb\_checkpoint...

Assignment2\_CaesarCypher\_SimpleMath&Statistics/.ipynb\_checkpoint...

Assignment2\_CaesarCypher\_SimpleMath&Statistics/.ipynb\_checkpoint...

Assignment2\_CaesarCypher\_SimpleMath&Statistics/.ipynb\_checkpoint...

CANCEL


5. Click on the folder icon and wait a few seconds for the runtime to connect.

5\_Decision Tree Classifiers.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

RAM Disk Editing

 A Notebook to Use Decision Tree Classifiers

This notebook shows how to train a decision tree to classify unseen instances.

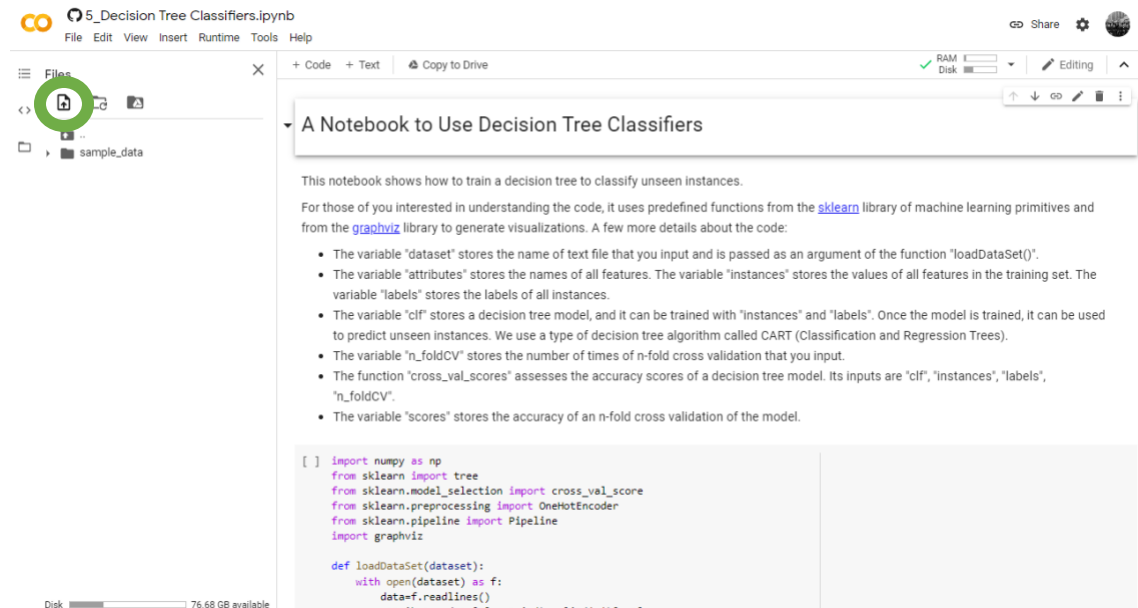
For those of you interested in understanding the code, it uses predefined functions from the [sklearn](#) library of machine learning primitives and from the [graphviz](#) library to generate visualizations. A few more details about the code:

- The variable "dataset" stores the name of text file that you input and is passed as an argument of the function "loadDataSet()".
- The variable "attributes" stores the names of all features. The variable "instances" stores the values of all features in the training set. The variable "labels" stores the labels of all instances.
- The variable "clf" stores a decision tree model, and it can be trained with "instances" and "labels". Once the model is trained, it can be used to predict unseen instances. We use a type of decision tree algorithm called CART (Classification and Regression Trees).
- The variable "n\_foldCV" stores the number of times of n-fold cross validation that you input.
- The function "cross\_val\_scores" assesses the accuracy scores of a decision tree model. Its inputs are "clf", "instances", "labels", "n\_foldCV".
- The variable "scores" stores the accuracy of an n-fold cross validation of the model.

```
[ ] import numpy as np
from sklearn import tree
from sklearn.model_selection import cross_val_score
from sklearn.preprocessing import OneHotEncoder
from sklearn.pipeline import Pipeline
import graphviz

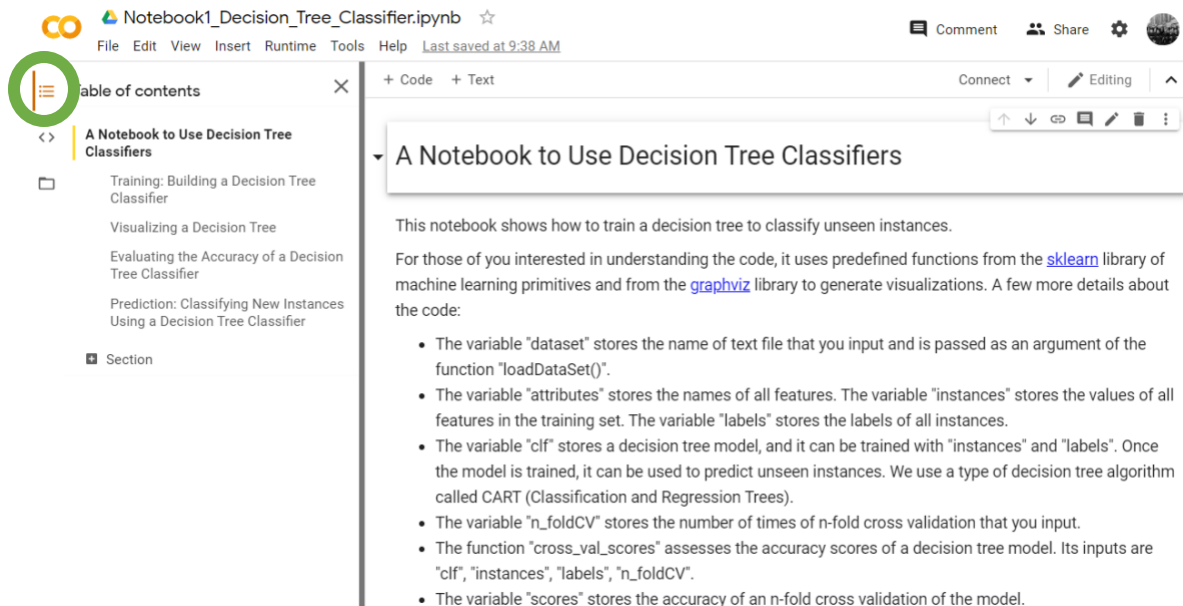
def loadDataSet(dataset):
    with open(dataset) as f:
        data=f.readlines()
```

6. The sample data folder contains datasets populated automatically by Google Colab. **DO NOT USE FOR YOUR HOMEWORK.**
- Some datasets are included as examples. To access them and download them into your Google Colab directory, run the corresponding cells in your notebook.
  - If you want to upload a file from your local computer, click on 'Upload to session storage' in the menu bar.



**Note:** Uploaded files are deleted after 12 hours. You might need to upload it again to use the file after 12 hours.

7. You can click on the Contents button to view and access any part of the notebook.



8. Use the notebooks to answer the questions given in the homework assignment.