



Deep Learning for Genomic Prediction



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Agenda & Introduction

Who are we?



María Inés
Fariello



Federico
Lecumberry



Graciana
Castro



Farielberry Lab



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



CICADA.uy

Centro Interdisciplinario en Ciencia de Datos y
Aprendizaje Automático



Resumen y objetivos del curso

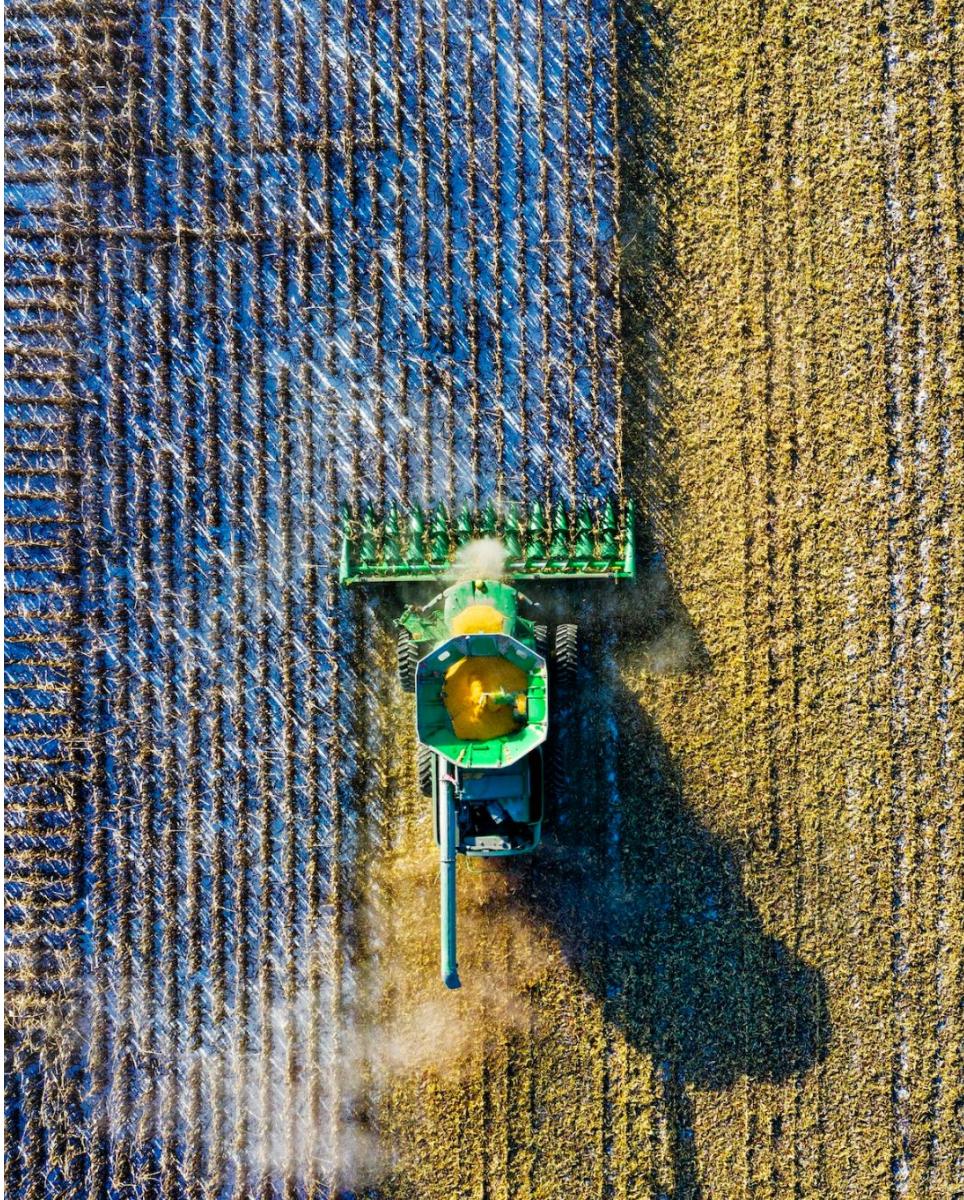


Se introducirán los conceptos principales del Aprendizaje Profundo (DL, Deep Learning) para lo que se cubrirán conceptos de Aprendizaje Automático (ML, Machine Learning) con énfasis en su aplicación a la Predicción Genómica (PG). También se utilizarán otros dominios como la visión artificial o el lenguaje natural para presentar ejemplos ilustrativos.

Resumen y objetivos del curso



Al finalizar el curso el estudiante comprenderá los fundamentos matemáticos del Aprendizaje Profundo, tendrá experiencia en la implementación de redes neuronales en Python y Pytorch, y del análisis de sus resultados.



Temario

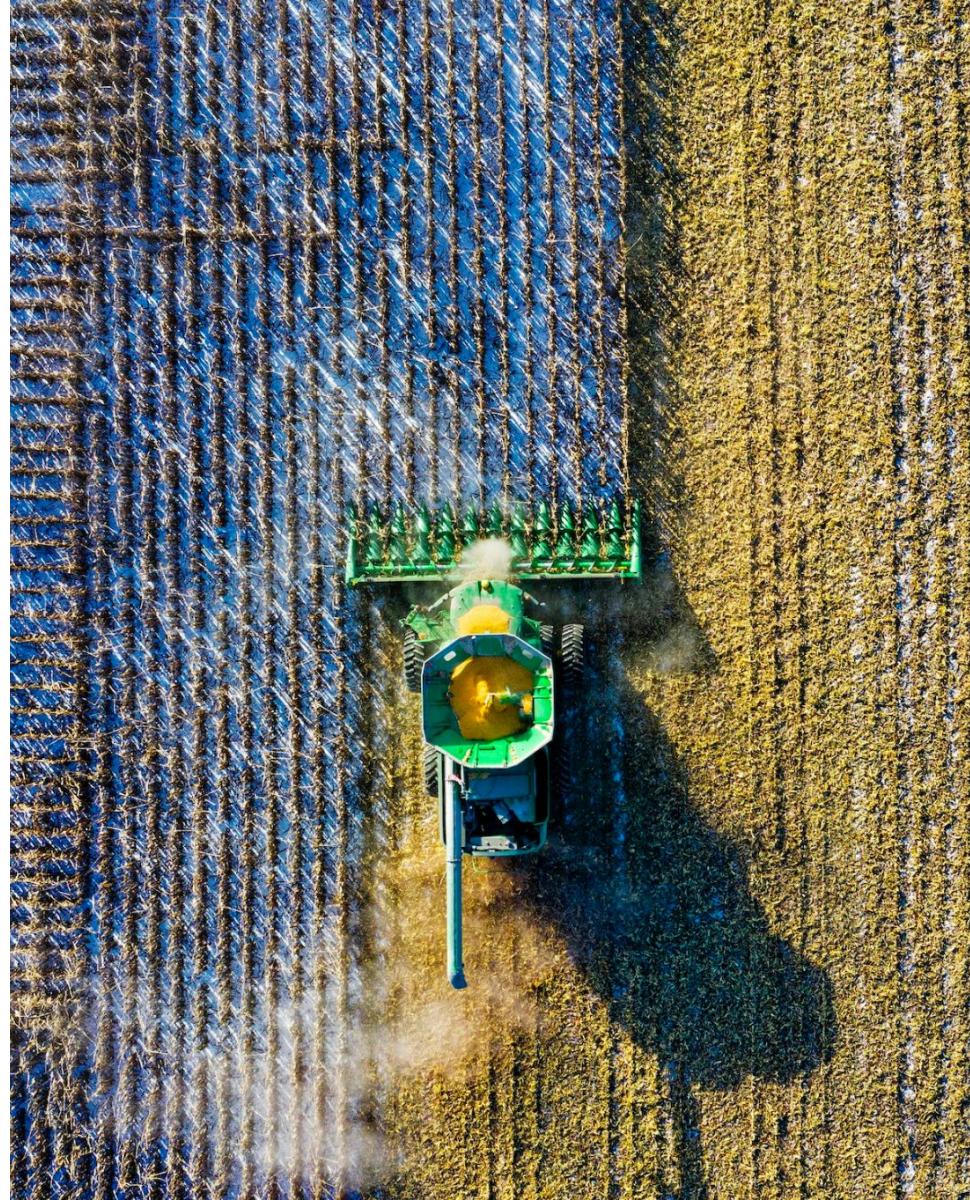
- • • • •
- 1. Intro to Genomic Prediction
- 2. Intro to Machine/Deep Learning
- 3. Convolutional Neural Networks
- 4. Graph Neural Networks
- 5. Transformers

Lunes a viernes de 9:00 a 13:00.



Evaluación

A conversar...

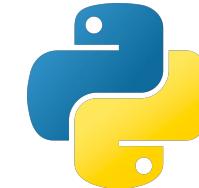




Environment setup for practicals

Environment setup

- Jupyter Notebook
- Google Colaboratory
- Python (numpy, matplotlib, altair, ...)
- PyTorch



Environment setup

1. Go to <https://classroom.google.com/>
2. Join with code **h6gl56t** clicking the +

The image consists of two vertically stacked screenshots of a web browser displaying the Google Classroom homepage. The top screenshot shows the main classroom feed with a floating menu. The 'Join class' button in this menu is circled in red. The bottom screenshot shows a modal dialog titled 'Join class'. It displays the user's account information ('You're currently signed in as Farielberry Lab, farielberry@gmail.com') and a 'Switch account' button. Below this, there is a 'Class code' section with the placeholder 'Ask your teacher for the class code, then enter it here.' A large input field contains the class code 'h6gl56t', which is also circled in red. To the right of the input field, the class code is displayed in a large, stylized blue font.

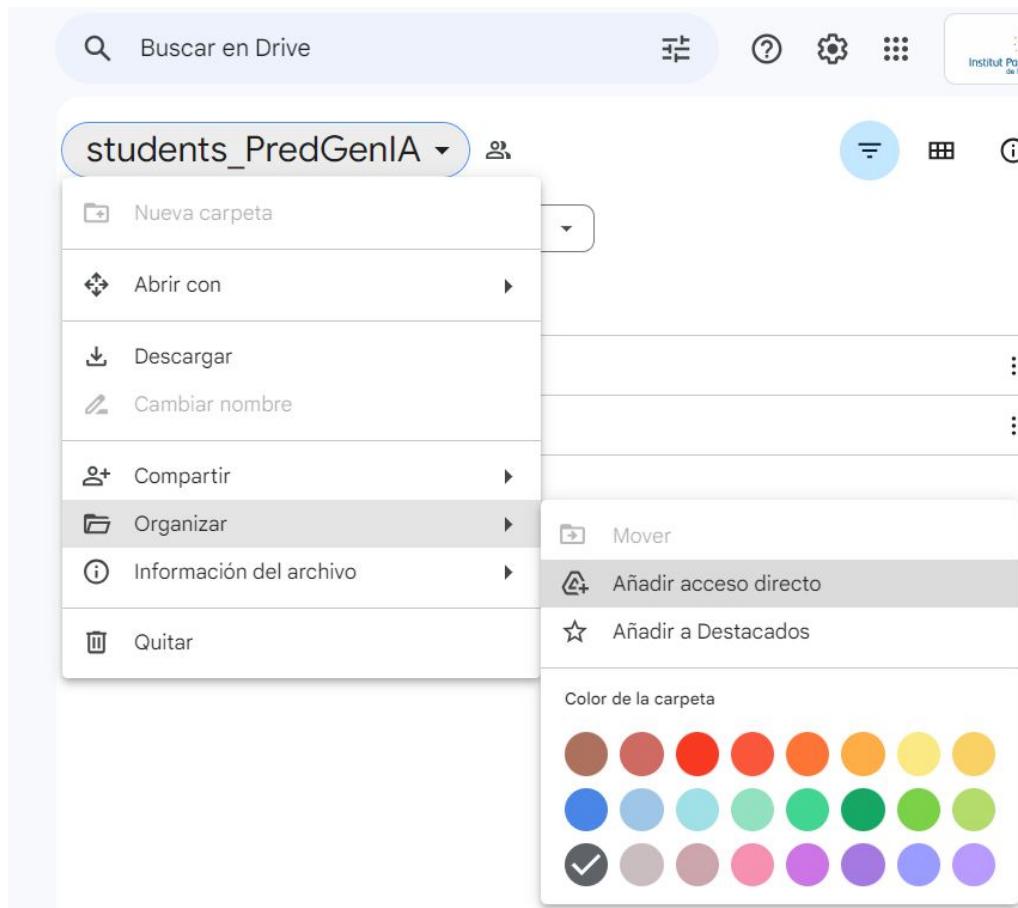
Join class

Join class

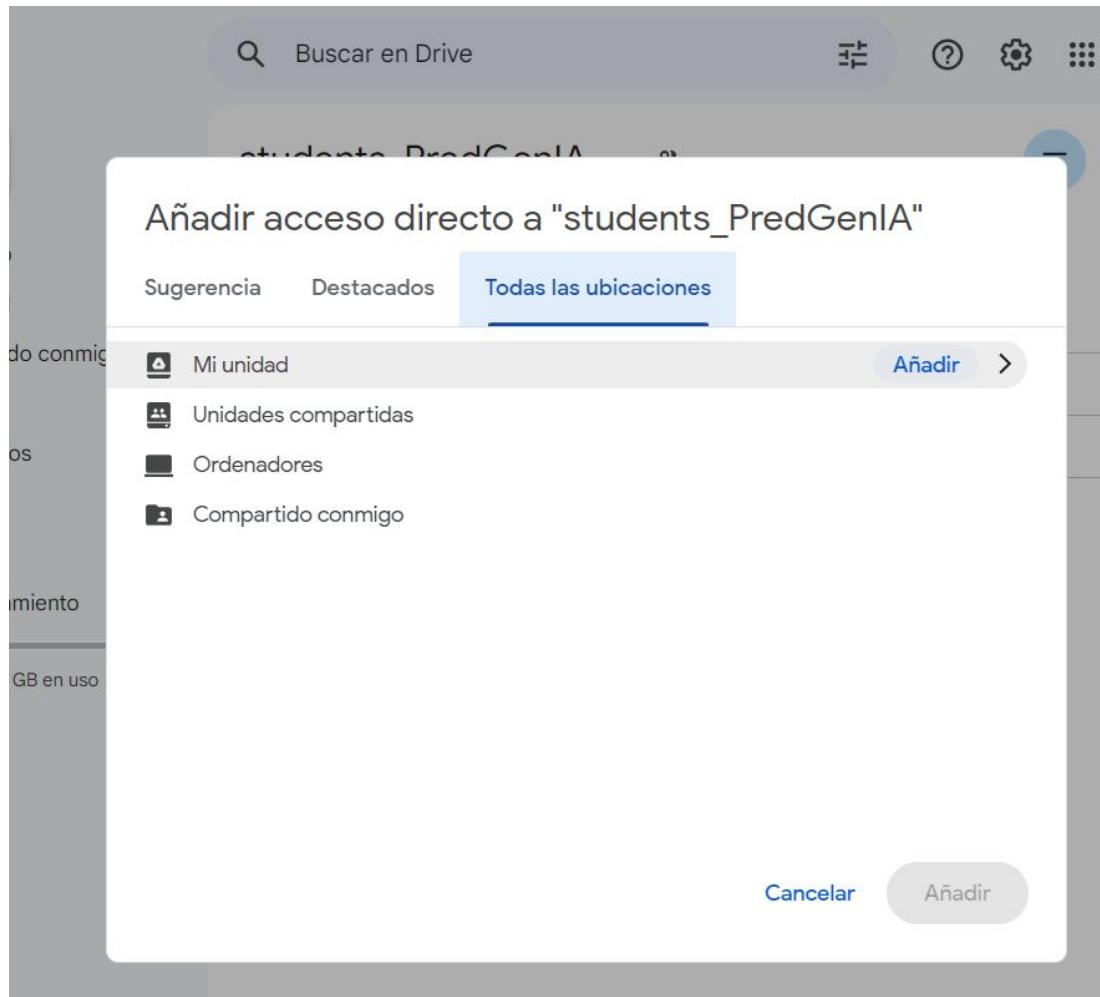
Join

h6gl56t

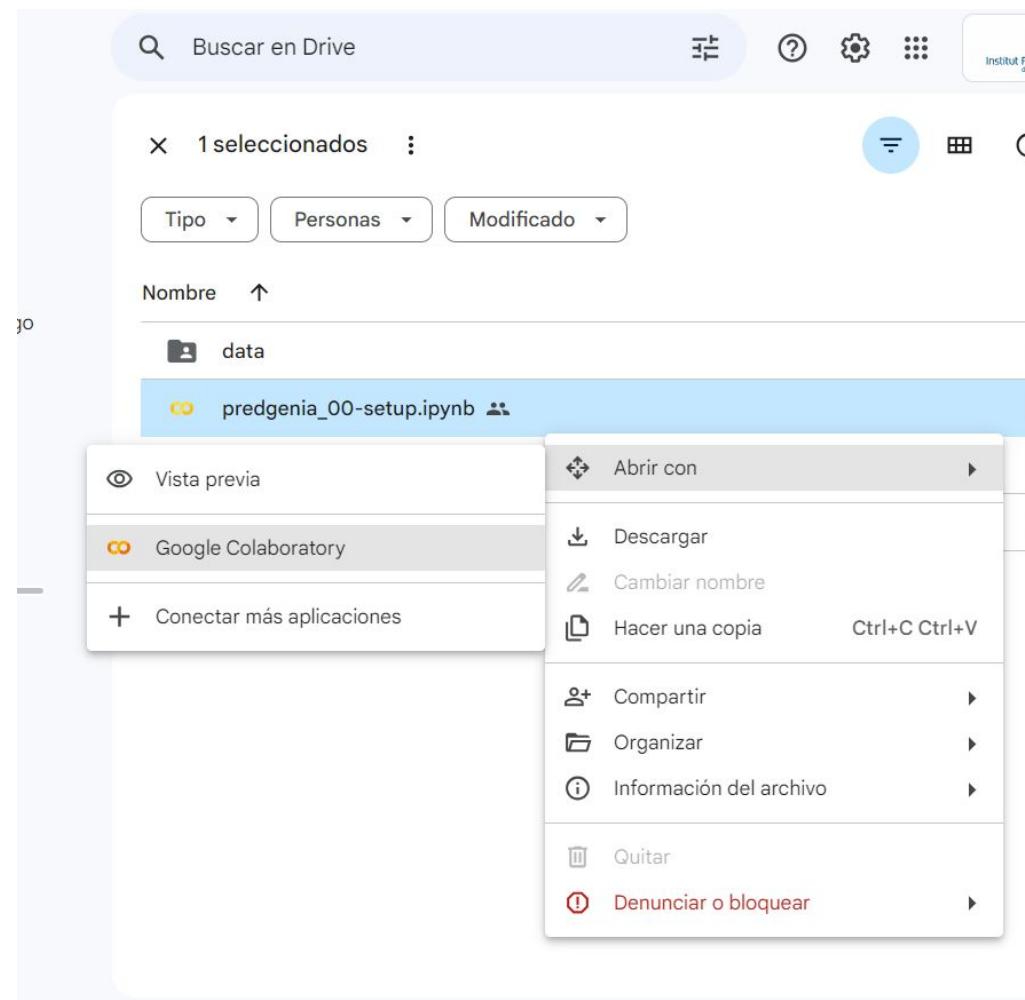
Environment setup



Environment setup



Environment setup



Environment setup

The screenshot shows a Jupyter Notebook interface with the following details:

- Title Bar:** predgenia_00-setup.ipynb
- File Menu:** File, Edit, View, Insert, Runtime, Tools, Help. A message "Changes will not be saved" is displayed.
- Submenu (opened at File):**
 - Locate in Drive
 - New notebook
 - Open notebook
 - Upload notebook
 - Rename
 - Move
 - Move to trash
 - Save a copy in Drive** (highlighted with a red box)
 - Save a copy as a GitHub Gist
 - Save a copy in GitHub
 - Save
 - Save and pin revision
 - Revision history
 - Download
 - Print
- Toolbar:** Share, Settings, Google Groups icon.
- Main Area:** A large text cell containing the title "for Genomic Prediction".
- Code Cell:** A code cell with the following content:

```
[ ] # It must end in '/'
path_to_predgenia = '/content/drive/MyDrive/Courses/PredGenIA/'
```

Find path to PredGenIA folder in your Google Drive in the left sidebar.

```
[ ] # It must end in '/'
path_to_predgenia = '/content/drive/MyDrive/Courses/PredGenIA/'
```

Environment setup

The screenshot shows a Jupyter Notebook interface with the following details:

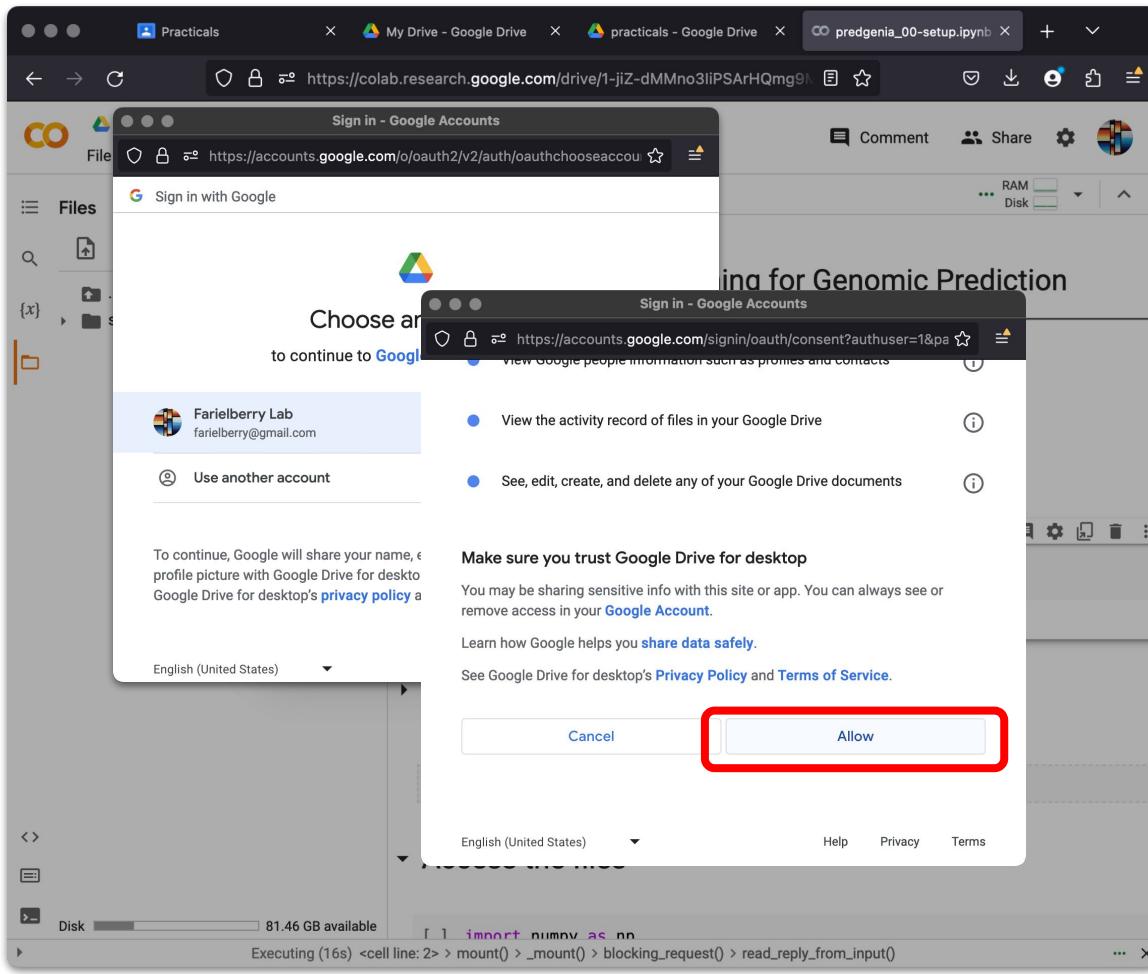
- Title Bar:** Shows the title "Copy of predgenia_00-setup.ipynb" and a star icon. The menu bar includes File, Edit, View, Insert, Runtime, Tools, Help, and a timestamp "Last saved at 5:55PM".
- Toolbar:** Includes Comment, Share, Settings, and a Google Groups icon.
- Code/Text Buttons:** + Code and + Text.
- Connect Button:** Connect dropdown and a file icon.
- Search and Filter:** A magnifying glass icon and a dropdown menu with an 'x' icon.
- Section Headers:** "PredGenIA: Deep Learning for Genomic Prediction" and "00: Environment setup".
- Section 1: 1. Mount Google Drive**
 - Description: "It will prompt asking for access to your Google Account."
 - Code: []

```
from google.colab import drive  
drive.mount('/content/drive')
```

Mounted at /content/drive
- Section 2: 2. PredGenIA folders setup**
 - Description: "Find path to PredGenIA folder in your Google Drive in the left sidebar."
 - Code: []

```
# It must end in '/'  
path_to_predgenia = '/content/drive/MyDrive/Courses/PredGenIA/'
```

Environment setup



Environment setup

The screenshot shows a Google Colab notebook titled "Copy of predgenia_00-setup.ipynb". The sidebar on the left displays a file tree. A red box highlights the folder "students_PredGenIA" under "MyDrive/Courses/PredGenIA". The main area contains two sections: "00: Environment setup" and "1. Mount Google Drive". The "1. Mount Google Drive" section includes a code cell that mounts the drive and a note about prompting for Google Account access. The "2. PredGenIA folders setup" section provides instructions to find the path to the PredGenIA folder in Google Drive and includes a code cell for setting the path variable.

File Edit View Insert Runtime Tools Help All changes saved

Comment Share G

Files

[x] bin
boot
content
drive
MyDrive
Colab Notebooks
students_PredGenIA
practicals
slides
2021.09.19.460980v1.f...
Info útil.gdoc
Notebooks en Nahual....
Pruebas semana 28 8....
URU1_fgene-12-73319...
Shareddrives
sample_data
datalab
dev
etc
home
lib
lib32

+ Code + Text

RAM Disk

PredGenIA: Deep Learning for Genomic Predict

00: Environment setup

1. Mount Google Drive

It will prompt asking for access to your Google Account.

```
[22s] [1] from google.colab import drive  
drive.mount('/content/drive')
```

Mounted at /content/drive

2. PredGenIA folders setup

Find path to PredGenIA folder in your Google Drive in the left sidebar.

```
[ ] # It must end in '/'  
path_to_predgenia = '/content/drive/MyDrive/Courses/PredGenIA/'
```

Environment setup

The screenshot shows a Google Colab notebook titled "Copy of predgenia_00-setup.ipynb". The sidebar on the left displays a file tree with a context menu open over the "MyDrive/Colab Notebooks" folder. The menu options include "Upload", "New file", "New folder", "Rename folder", "Delete folder", and "Copy path", with "Copy path" highlighted by a red rectangle. The main content area shows two sections: "00: Environment setup" and "1. Mount Google Drive". The "Mount Google Drive" section contains code to import the `drive` module and mount the drive at `/content/drive`, followed by a success message. The next section, "2. PredGenIA folders setup", instructs users to find the path to the PredGenIA folder in their Google Drive. A code cell shows the assignment of `path_to_predgenia` to the mounted drive path.

Copy of predgenia_00-setup.ipynb

All changes saved

Comment Share G

RAM Disk

Files

bin
boot
content
drive
MyDrive
Colab Notebooks

Upload
New file
New folder
Rename folder
Delete folder
Copy path
Refresh

00: Environment setup

1. Mount Google Drive

It will prompt asking for access to your Google Account.

```
[22s] [1] from google.colab import drive  
drive.mount('/content/drive')
```

Mounted at /content/drive

2. PredGenIA folders setup

Find path to PredGenIA folder in your Google Drive in the left sidebar.

```
[ ] # It must end in '/'  
path_to_predgenia = '/content/drive/MyDrive/Courses/PredGenIA/'
```

Environment setup

Copy of predgenia_00-setup.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share G

RAM Disk

Files

{x} bin boot content drive MyDrive Colab Not... students_... practicals slides 2021.09.19... Info útil.g... Notebooks... Pruebas se... URU1_fgen... Shareddrives sample_data datalab dev etc home lib lib32

+ Code + Text

22s [1] `from google.colab import drive`
[1] `drive.mount('/content/drive')`
Mounted at /content/drive

2. PredGenIA folders setup

Find path to PredGenIA folder in your Google Drive in the left sidebar.

[] # It must end in '/'
path_to_predgenia = '/content/drive/MyDrive/students_PredGenIA/' ← Paste

3. Check files

[] `import numpy as np`
`# Define path to data.`
`path_to_phenotype_data_file = path_to_predgenia + 'practicals/data/wheat_phenotype.csv'`
`path_to_genotype_data_file = path_to_predgenia + 'practicals/data/wheat_genotype.csv'`
`# Load the files`
`X = np.load(path_to_genotype_data_file)`
`y = np.load(path_to_phenotype_data_file)`
`# Print some information.`
`print('Individuals: %d\nGenotype size: %d' % (np.shape(X)[0], np.shape(X)[1]))`