



# Final Project

# Building a ML

# wrapper

2021-22 Course

# Objectives

- Main Objective: Design a programming language to mimic the workflow of a Machine Learning (ML) task.
- Subobjectives:
  - Design a programming language.
  - Design a grammar to represent the previous language.
  - Design a lexical analyzer (lexer) for the language.
  - Design a syntactic analyzer (parser) for the language.
- Up to three students *can* develop the Project in group.

# Submission

- Upload a zip file containing:
  - Memory with all considerations during design process
  - User manual
  - SLY files (Lexer and Parser)
  - Working example

Deadline:

5th of June

# Evaluation

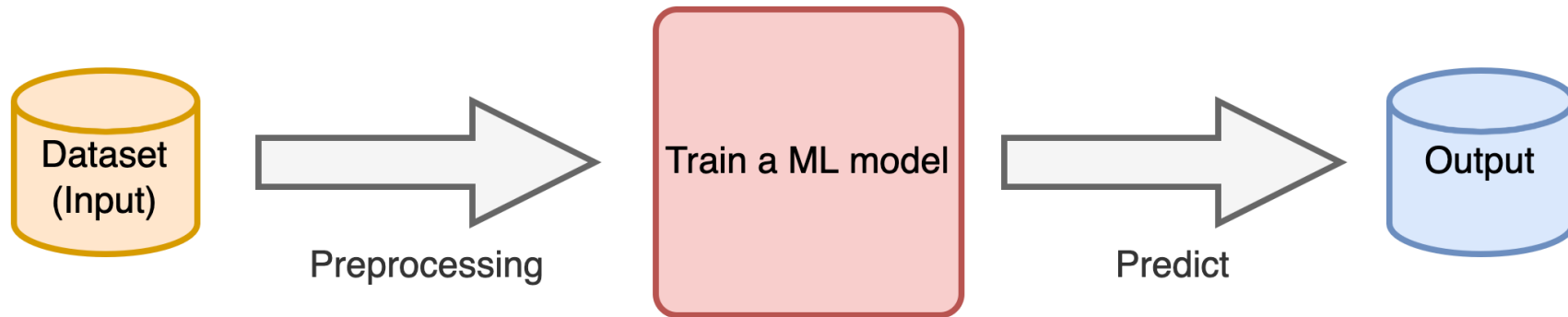
Concepto	100%	75%	25%	0%
Lenguaje diseñado (2.5pts)	Es fácil de usar, de aprender, y muy descriptivo	Es muy descriptivo, pero no es sencillo de utilizar	Es sencillo de utilizar, pero no muy descriptivo	No es lo suficientemente descriptivo
Gramática (2.5pts)	Genera el lenguaje diseñado y es correcta	Genera el lenguaje diseñado, pero presenta ambigüedades	Algunas expresiones del lenguaje no se pueden generar	Muchas expresiones del lenguaje no se pueden generar
Implementación de la gramática en SLY (2.5pts)	Se corresponde con la gramática diseñada. Sin conflictos ni errores	Implementación correcta, pero se ayuda de herramientas de SLY para evitar conflictos	Se corresponde con la gramática diseñada. Hay pocos conflictos, pero sin errores	No se corresponde con la gramática diseñada, hay muchos conflictos y/o errores
Manual de usuario (1.5pts)	Completo y bien descrito	Completo, pero no está bien descrito	Faltan algunas cosas	Faltan muchas cosas
Memoria (1pt)	Cuidadosamente presentada. Todo bien explicado	Bien presentada y explicado	Presentación mejorable o faltan cosas por explicar	Presentación mejorable y faltan cosas por explicar



# ML Models



# Workflow



# Preprocessing

- Certain datasets will be selectable by the grammar. These will be provided to you in the same manner as the Pattern Library, but in a format easily readable by Python.

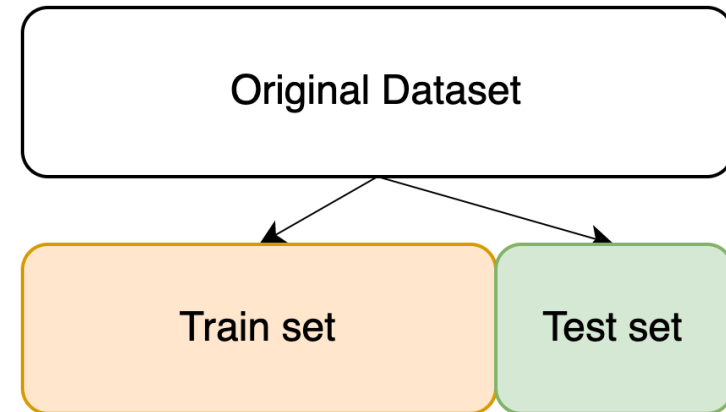
```
pd.read_csv('Iris.csv') -> Matrix representing Iris dataset
```

Datasets = {Iris, Image, Pima, Wine}

# Preprocessing

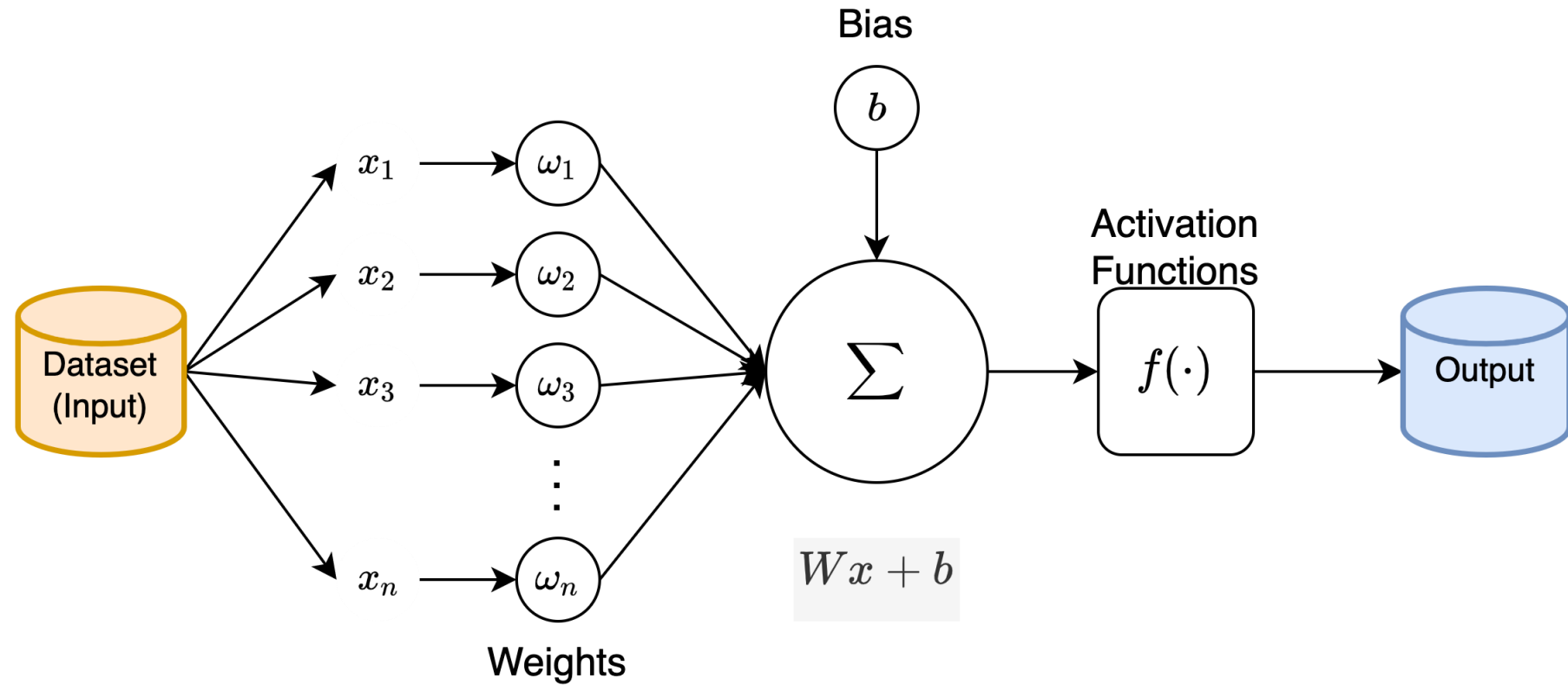
- This step can have any of this operations:
  - Change values of a dataset -> Normalization
  - Split dataset into train and test sets (Code to do it will be provided)

$$\frac{x - x_{min}}{x_{max} - x_{min}}$$





# Neural Networks



# Neural Networks

- Are we go to implement NNs? The answer is no, it is not in the scope of this subject. We are going to use Keras library.

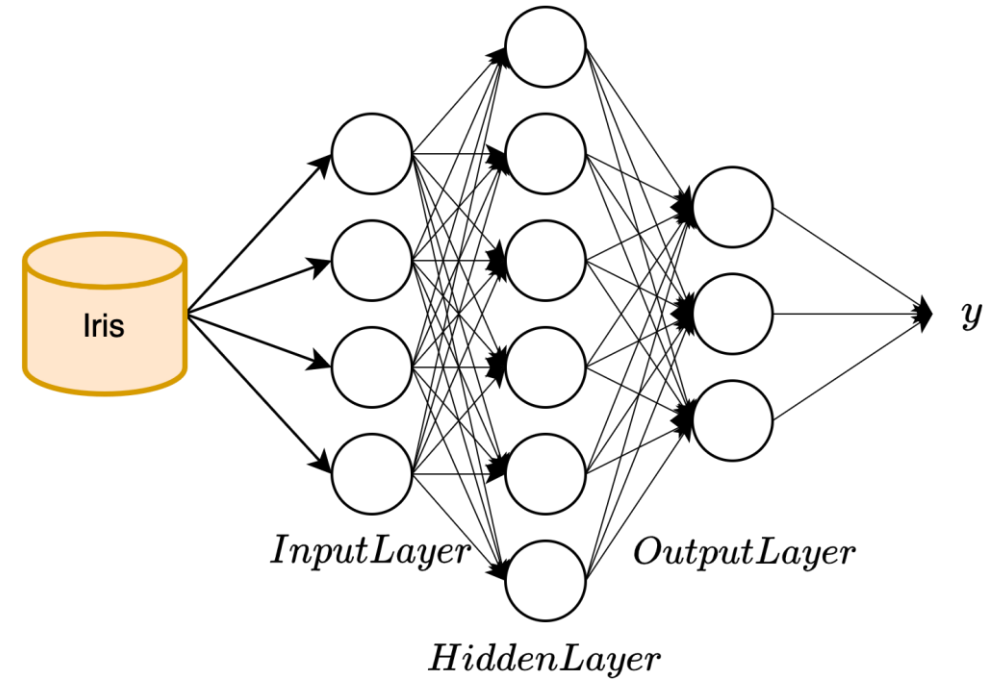
In essence, the language could be seen as a wrapper for this library. It is easy to use, and we can get rid of the topics out of this subject.



# Keras Library

- Keras usage is straightforward: Just add layers with output sizes.

```
model = Keras.models.Sequential( )  
model.add(Dense(6, 'relu'))  
model.add(Dense(3, 'softmax'))  
model.compile(loss='categorical_crossentropy',  
              metrics=['accuracy'])
```



# Keras Library

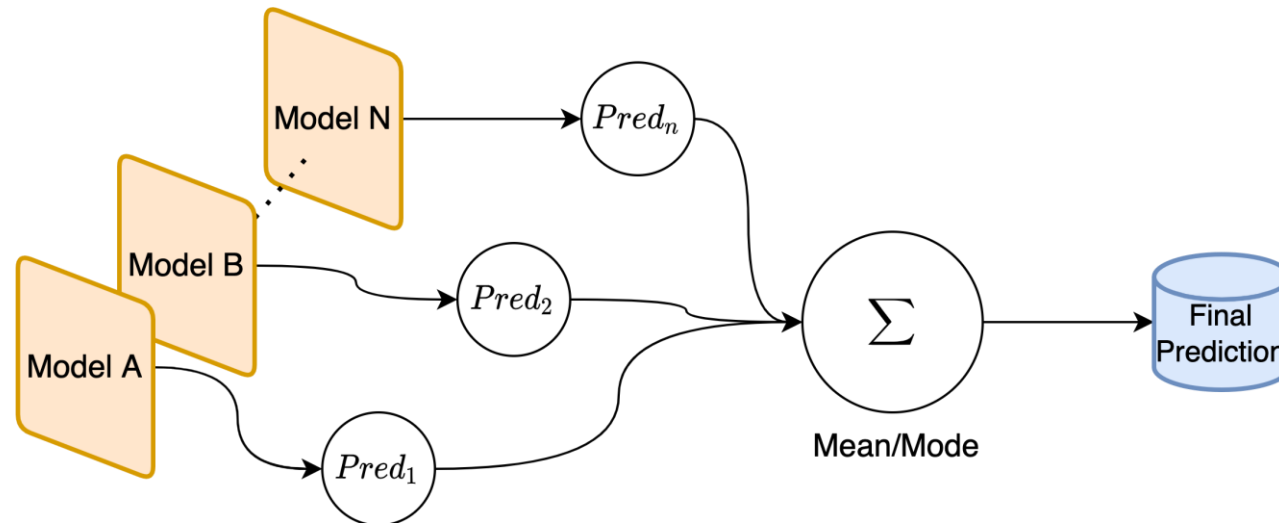
- Notice that we only need:
  - How many layers the model will have, at least one.
  - Each layer is composed of numbers of neurons for the output and the activation function. Last one must be 'softmax'.
- Then, we can use the model:
  - `model.fit(X, y, epochs=200)` # Iters can be seen somehow as the iterations.
  - `model.predict(nuevaX)`

# Keras Library

- Further details about usage can be omitted. In virtual campus a document will be provided in order to cover some details and offer you fragments of code.
- Remember, the scope of this practice is to design a grammar!

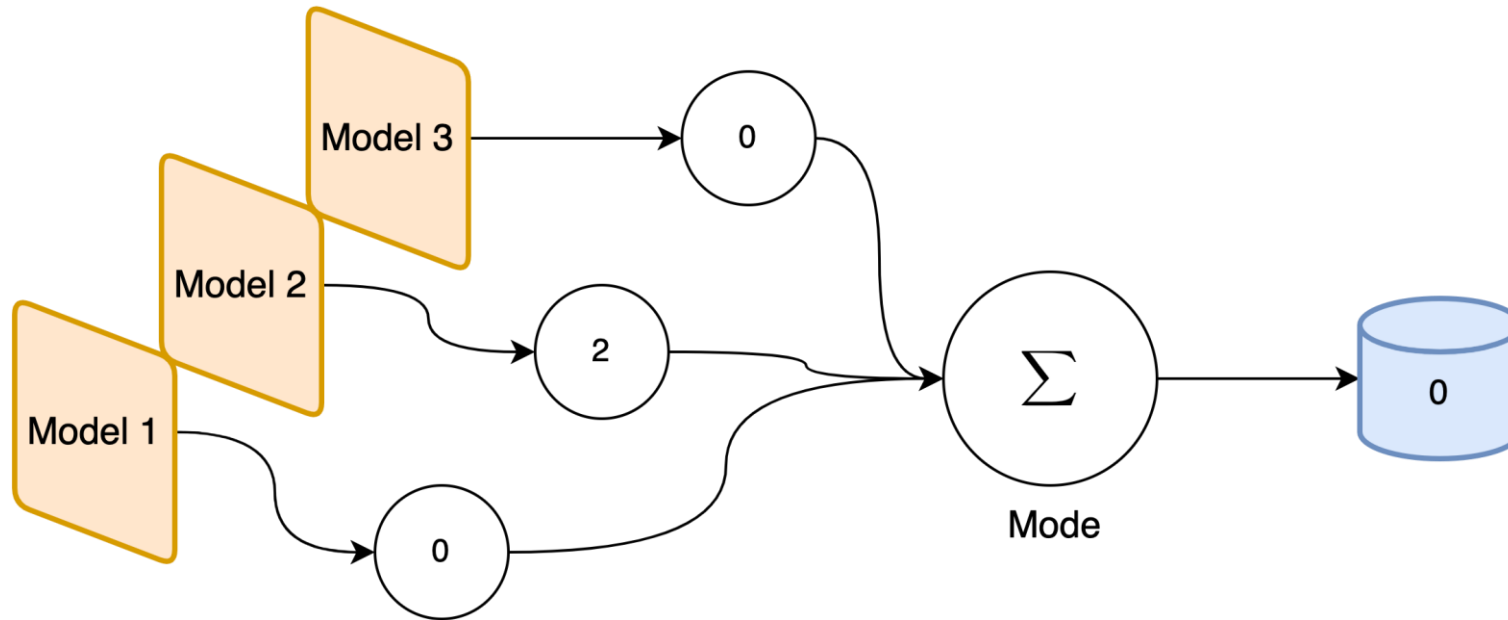
# Ensembling

- It is based on the principle of “There is strength in numbers”.
- The same applies in the context of ML.
- There are several approaches but, we are going to use the easiest.



# Ensembling

- In Keras context we just need to perform the predict function using each model and then use the mean or the mode.



# Summary

- Design a grammar capable of:
  - Apply certain preprocessing operations such as splitting dataset or scale the dataset (normalization).
  - Create a MLP model just by adding layers with certain parameters.
  - Being able to combine different models output into a final predictions.
  - Selecting the dataset that will be used in the process.

*If you have any doubt, don't hesitate and ask!*

