

# Notes Hackathon

X-AI 2022

October 2022

## 1 Introduction

Ces notes donnent quelques éléments afin de vous aider au cours du Hackathon.

## 2 Rappel de base de machine learning

Un problème de machine learning supervisé se présente sous la forme suivante. Étant donné un dataset  $\mathcal{D} = (x_i, y_i)_{i \in [1, N]}$  où  $(x_i)_{i \in [1, N]}$  représente les exemples et  $(y_i)_{i \in [1, N]}$  les labels correspondants, on cherche une fonction  $f$  dépendant d'un paramètre  $\theta$  qui minimise l'erreur entre les prédictions de notre modèle modélisé par  $f_\theta(x_i)$  et les labels  $y_i$ .

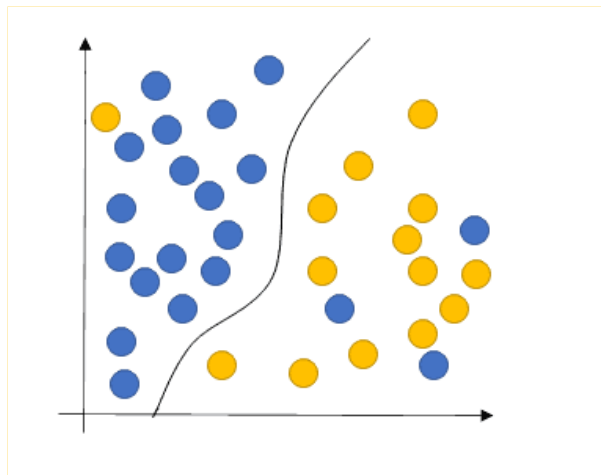


Figure 1: Exemple de problème de classification

La première étape de votre projet sera donc de déterminer ces éléments afin de choisir le modèle et la fonction mesurant l'erreur qui est la plus appropriée.

## 3 Implémentation

La plupart des projets de machine learning sont implémentés en **Python**, nous vous imposons donc d'utiliser ce langage de programmation afin de concevoir vos modèles.

Afin de pouvoir exécuter du code Python facilement l'utilisation de notebook est fortement recommandée. Pour cela, utiliser [Google Colab](#) est conseillé, il permet partager votre code (mais pas de le modifier à plusieurs simultanément). Il vous permet par ailleurs d'accéder à des GPUs (carte graphiques) qui vous permettront d'entraîner votre modèle

plus rapidement. Pour cela, il faut aller dans l'onglet "Modifier" puis "Paramètres du notebook", faire dérouler le menu "Accélérateur matériel" et sélectionner "GPU". Il vous suffit de relancer votre notebook depuis le début et vous disposerez du GPU pour faire vos calculs.

## 4 Packages suggérés

Afin d'éviter de réinventer la roue, vous pouvez bien entendu utiliser des packages Python existants. Nous recommandons en particulier [Tensorflow](#) ou [PyTorch](#) afin d'implémenter votre modèle. Vous trouverez facilement de nombreuses informations sur comment implémenter un modèle en utilisant le package de votre choix.

**Il ne sera évidemment pas toléré que vous utilisiez des packages comme OpenCV qui comporte déjà des modèles pré-entraînés afin de résoudre le problème du Hackathon.**

## 5 Erreur à éviter

Au cours de ce projet vous allez malencontreusement être face à de nombreux bugs. Si vous choisissez d'utiliser Google Colab, vous allez devoir importer le dataset sur un Google Drive puis y accéder directement depuis le notebook. Il est recommandé d'importer le dataset directement en format compressé ".zip" pour éviter l'apparition de doublons de fichiers lors de l'import dans le Drive.

## 6 Application Web

Pour la partie présentation au client, on suggère dans un premier temps d'utiliser le package Python [Streamlit](#) (très facile à utiliser). Pour un travail plus ambitieux, on pourra s'intéresser à Flask.

Bonne chance pour le Hackathon !  
La team du Binet IA 2020