

Mini-Projet d'Apprentissage Supervisé Régression Linéaire et Régression Logistique

Master 1 Intelligence Artificielle (M1 IA) Année universitaire : 2025–2026

Étudiant : Mohamed Mahmoud Said

Matricule : C34666

1. Introduction

L'apprentissage supervisé est une branche fondamentale du Machine Learning qui consiste à entraîner un modèle à partir de données étiquetées afin de prédire une variable cible. L'objectif de ce mini-projet est de consolider les concepts théoriques vus en cours à travers l'implémentation pratique de deux modèles essentiels : la régression linéaire et la régression logistique. Ce travail permet de comprendre les différentes étapes d'un pipeline de Machine Learning : préparation des données, modélisation, entraînement, évaluation et interprétation des résultats.

2. Régression Linéaire

2.1 Présentation des données

Un jeu de données de type régression a été utilisé, contenant plusieurs variables explicatives et une variable cible numérique. Une analyse exploratoire a été réalisée afin de mieux comprendre les relations entre les variables.

2.2 Analyse des corrélations

Une matrice de corrélation a été construite et visualisée à l'aide d'une heatmap afin d'identifier les relations importantes entre les variables et la variable cible.

2.3 Modélisation mathématique

Le modèle est défini par l'équation : $y = \beta_0 + \sum \beta_i x_i + \epsilon$. Il a été entraîné à l'aide de la bibliothèque scikit-learn.

2.4 Évaluation du modèle

Les performances ont été évaluées à l'aide du coefficient de détermination R^2 et de l'erreur quadratique moyenne (MSE). Les résultats obtenus montrent une bonne capacité de prédiction du modèle.

2.5 Interprétation des coefficients

L'analyse des coefficients permet d'identifier l'importance de chaque variable explicative dans la prédiction finale.

3. Régression Logistique

3.1 Présentation des données

Le dataset Iris a été utilisé et transformé en un problème de classification binaire.

3.2 Préparation des données

Les données ont été normalisées et divisées en ensembles d'entraînement et de test afin d'améliorer la performance du modèle.

3.3 Modélisation

La probabilité d'appartenance à une classe est modélisée à l'aide de la fonction sigmoïde :

$$[P(y=1|x) = \frac{1}{1 + e^{-z}}]$$

3.4 Évaluation du modèle

Le modèle a été évalué à l'aide d'une matrice de confusion, ainsi que des métriques Accuracy, Précision et Recall. Les résultats montrent une bonne capacité de classification.

4. Conclusion

Ce mini-projet a permis de mettre en pratique les bases de l'apprentissage supervisé. La régression linéaire et la régression logistique se sont révélées efficaces pour les tâches de prédiction et de classification. Ce travail constitue une base solide pour l'étude de modèles plus avancés en Machine Learning.

5. Outils et Technologies

Python, NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, Google Colab, GitHub.