Rapport Extension Mini Projet
Machine Learning

Alexis Araujo

Introduction

Ce projet avait pour objectif de construire un modèle de classification entre les tulipes et les roses en se basant sur des images de ces fleurs. J'ai utilisé un algorithme d'apprentissage supervisé : le perceptron, afin de prédire si une image donnée représente une tulipe ou une rose en se basant sur leurs caractéristiques visuelles.

Création du Dataset

J'ai commencé par créer un ensemble de données en utilisant des images représentant des tulipes et des roses. Chaque image a été redimensionnée à une taille spécifique (64x64 pixels) pour assurer la cohérence des données.

Pour chaque catégorie (tulipe et rose), j'ai extrait les images png, les ai redimensionnées et converties en tableaux numpy. J'ai ensuite créé les étiquettes correspondantes : 0 pour les roses et 1 pour les tulipes. Cela a permis de construire le dataset avec les images et les étiquettes correspondantes pour l'entraînement et les tests. Tout a été fait dans le fonction load_data_train() et load_data_test() qui vont chercher le dossier correspondant à l'entraînement et aux tests.

De plus, j'ai pu réaliser un test via uniquement le dossier d'entraînement en ajoutant un split pour partitionner l'entraînement et les tests avec des données choisis aléatoirement (le code a été mis en commentaire).

Apprentissage avec Perceptron

Après avoir chargé et prétraité les données d'entraînement et de test, j'ai utilisé un perceptron pour l'apprentissage. Voici les étapes clés de cet apprentissage :

Paramétrage et visualisation des données

J’ai réalisé des affichages pour connaître le nombre de données d'entraînement et de test, ainsi que la dimension 64,64 et le 3 qui signifie des données en RVB.

J’ai ajouté une ligne de reshape avec la 1er colonne du tableau pour passer à une classification binaire.

Un affichage avec np.unique pour observer la répartition de la classification des données de 1 et de 0.

Pour visualiser les images j’ai utilisé la librairie matplotlib, en réalisant une boucle sur les images d'entraînement.

Ensuite je normalise les données pour avoir une bonne forme.

Initialisation des paramètres

Les poids (W) et le biais (b) du modèle ont été initialisés aléatoirement pour démarrer le processus d'apprentissage.

Modèle de prédiction

Le modèle a été construit pour calculer les prédictions en utilisant une fonction sigmoïde pour déterminer si une image donnée représente une tulipe ou une rose.

Calcul de la perte

J'ai utilisé la fonction de perte logistique pour évaluer les performances du modèle.

Calcul des gradients

Les gradients pour les poids et le biais ont été calculés pour ajuster ces paramètres lors de l'apprentissage.

Mise à jour des paramètres

J'ai mis à jour les poids et le biais en utilisant la descente de gradient afin d'améliorer progressivement la performance du modèle.

Entraînement du modèle

L'apprentissage s'est déroulé sur un nombre d'itérations fixées à 15000. À chaque itération, le modèle a calculé les prédictions, les gradients, mis à jour les poids et le biais, tout en enregistrant la perte pour évaluer les performances du modèle.

Évaluation et Visualisation

J'ai évalué les performances du modèle en traçant les courbes de coût et d'exactitude pour les ensembles d'entraînement et de test. Ces courbes ont montré la

progression de l'apprentissage et l'ajustement du modèle au fil des itérations. Il y a cependant un problème d’overfiting (de surapprentissage) que je n’arrive pas à régler.

Prédiction d’une donnée

J’ai évalué sur une donnée d’image si le modèle fonctionnait. La conclusion est que le modèle est surentraîné et il a du mal à prédire les images qu’il n’a pas vues. C' est pourquoi il faut reprendre le lancement du programme plusieurs fois. Je n’ai pas réussi à régler ce problème.

Conclusion

Ce mini projet m'a permis de mettre en pratique l'apprentissage supervisé, en particulier l'utilisation du Perceptron pour la classification entre les tulipes et les roses en se basant sur des données visuelles. Malgré quelques ajustements, le modèle a démontré qu’il était surentraîné et a du mal à distinguer les deux types de fleurs.

En tout cas j’aimerai savoir d’où provient cette erreur si vous pouviez m’aider je serai reconnaissant, car je suis très intéressé par l’objet de ce mini-projet et du cours.