

Oh Flowers



ELFEKIH Ons

Date : 04 Décembre 023

Heure : 9h

Matière:PPP TI

Programme

Etude de l'existant

Objectif

Exigences Fonctionnelles

Le fonctionnement de KNN

Technologies utilisées

Contraintes Techniques

Conclusion

Etude de l'existant

- Il contient des mesures de 150 fleurs d'iris appartenant à trois espèces différentes :
Setosa, Versicolor et Virginica.
- Pour chaque fleur, quatre caractéristiques sont enregistrées :
 - Longueur des sépales (en centimètres)
 - Largeur des sépales (en centimètres)
 - Longueur des pétales (en centimètres)
 - Largeur des pétales (en centimètres)

Objectif

L'objectif principal est de former un modèle d'apprentissage automatique capable de classer avec précision les fleurs d'iris dans l'une des trois espèces en se basant sur ces quatre caractéristiques.

Exigences Fonctionnelles

Collecte de Données

Télécharger les données.

Prétraitement des Données :

Diviser les données en un ensemble d'apprentissage (70-80%) et un ensemble de test (30-20%).

Modèle de Classification : (K-NN).

K-Nearest Neighbors, un algorithme d'apprentissage automatique. Il fonctionne sur le principe de la proximité des données.

Entraînement du Modèle :

Appliquer le modèle sur l'ensemble d'apprentissage pour qu'il apprenne à classer les fleurs en fonction des caractéristiques.

Évaluation du Modèle :

Évaluer la performance du modèle en utilisant des métriques telles que l'exactitude, la précision, le rappel et le score F1.
Générer une matrice de confusion.

Fonctionnement de KNN

Collecte de données:

Collecter un ensemble de données sur les fleurs d'iris, y compris des mesures telles que la longueur et la largeur des pétales et des sépales.

Préparation des données:

divisez l'ensemble de données en ensembles d'apprentissage et de test.

Choix de la valeur de K:

Choisir une valeur de K appropriée, qui détermine le nombre de voisins à considérer pour la classification. Le choix de K est crucial et peut avoir un impact significatif sur les performances du modèle.

Entraînement du modèle:

Pour chaque point de données dans l'ensemble d'apprentissage, l'algorithme K-NN calcule la distance entre ce point et le point de données que vous souhaitez classer.

Les K points d'apprentissage les plus proches sont sélectionnés.

Fonctionnement de KNN

Classification:

l'algorithme examine les classes des K voisins les plus proches.

L'étiquette de classe la plus fréquente parmi ces K voisins devient la classe prédite pour l'iris avec des caractéristiques inconnues.

Validation:

évaluez les performances du modèle en utilisant l'ensemble de données de test.

Optimisation:

Ajuster la valeur de K , explorer différentes mesures de distance et effectuer d'autres ajustements pour améliorer les performances du modèle.

Prédiction:

Utiliser le modèle en cas de satisfaction, pour prédire la classe de nouvelles fleurs d'iris en fournissant leurs caractéristiques en entrée.

Technologies utilisées

Langage de Programmation:

Python

Bibliothèques d'Apprentissage Automatique :

NumPy: Pour calcul numérique

Pandas: Pour les structures des données

Bibliothèques de Visualisation:

Matplotlib: Pour les graphiques

Environnements de Développement :

Environnements Virtuels (virtualenv, Conda:gestionnaire de paquets)

IDE (PyCharm, Visual Studio Code, Google Colab)

Outils pour l'Analyse de Données :

Google Colab Notebooks

Validation et Optimisation du Modèle :

Validation Croisée: Pour évaluer la performance

Contraintes techniques

Utilisation de Python:

Le langage de programmation utilisé pour ce projet doit être Python. Cela signifie que le code source du modèle de classification et les outils associés doivent être écrits en Python.

Bibliothèques scikit-learn pour l'apprentissage automatique:

Scikit-learn offre une large gamme d'algorithmes d'apprentissage automatique, des outils de prétraitement des données et des fonctionnalités d'évaluation de modèles.

Compatibilité avec des systèmes courants :

Le modèle de classification que vous développez doit être conçu de manière à pouvoir fonctionner sur des systèmes courants. Cela implique de prendre en compte les spécifications matérielles et logicielles courantes pour garantir que le modèle soit utilisable sur des ordinateurs et des infrastructures couramment disponibles.



Conclusion

le problème de classification des fleurs d'Iris sert d'exemple fondamental de classification supervisée en apprentissage automatique, aidant les apprenants à comprendre les étapes clés pour construire et évaluer des modèles de classification.

Merci