Atividade Avançada: Árvores de Decisão, KNN e SVM em Python

Tiago Alves de Oliveira

Objetivo

Desenvolver uma árvore de decisão mais complexa com **10 perguntas** e explorar datasets do **Kaggle** e do **UCI Machine Learning Repository** para criar novas aplicações utilizando algoritmos de aprendizado supervisionado como **Árvore de Decisão**, **KNN (K-Nearest Neighbors)** e **SVM (Support Vector Machine)**.

Parte 1: Desafio - Árvore de Decisão com 10 Perguntas

Implemente uma árvore de decisão que ajude uma pessoa a decidir entre diferentes hobbies ou carreiras, com base nas respostas a 10 perguntas.

Exemplo de Estrutura de Perguntas

Aqui está uma sugestão de estrutura para as perguntas:

1. Você gosta de trabalhar em equipe? (Sim/Não)
2. Prefere atividades ao ar livre? (Sim/Não)
3. Gosta de lidar com números? (Sim/Não)
4. Prefere usar habilidades artísticas? (Sim/Não)
5. Se sente confortável trabalhando com tecnologia? (Sim/Não)
6. Gosta de resolver problemas complexos? (Sim/Não)
7. Prefere atividades que envolvam comunicação? (Sim/Não)
8. Se interessa por cuidar de outras pessoas? (Sim/Não)
9. Gosta de desafios físicos? (Sim/Não)
10. Prefere trabalhar em ambientes organizados? (Sim/Não)

Exemplo de Código Inicial

```
def arvore_decisao():
    equipe = input("Voc  gosta de trabalhar em equipe? (sim/n o): ").
        strip().lower()
    if equipe == "sim":
        ar_livre = input("Prefere atividades ao ar livre? (sim/n o): "
            ).strip().lower()
        if ar_livre == "sim":
            print("Sugest o: Explore atividades como esportes
                coletivos ou jardinagem.")
        else:
            tecnologia = input("Se sente confort vel trabalhando com
                tecnologia? (sim/n o): ").strip().lower()
            if tecnologia == "sim":
                print("Sugest o: Considere  reas  de tecnologia ou
                    programa  o.")
            else:
                print("Sugest o: Experimente hobbies criativos como
                    pintura ou m sica.")
    else:
        numeros = input("Gosta de lidar com n meros? (sim/n o): ").
            strip().lower()
        if numeros == "sim":
            print("Sugest o:  reas  como finan as ou engenharia podem
                ser interessantes.")
        else:
            print("Sugest o: Considere hobbies individuais como
                leitura ou fotografia.")

arvore_decisao()
```

Adapte esse exemplo para incluir mais perguntas e oferecer respostas personalizadas com base nas combinações de respostas.

Parte 2: Exploração de Datasets no Kaggle e UCI

Agora, vamos explorar dados reais para aplicar algoritmos de aprendizado supervisionado.

Instruções

1. Visite os seguintes sites:
   - **Kaggle**: `https://www.kaggle.com`
   - **UCI Machine Learning Repository**: `https://archive.ics.uci.edu/ml/index.php`
2. Pesquise datasets relacionados a temas de seu interesse, como:
   - Saúde (e.g., diagnósticos médicos)
   - Finanças (e.g., classificação de crédito)
   - Educação (e.g., desempenho acadêmico)

- Esportes (e.g., análise de desempenho de jogadores)

3. Faça o download de um dataset e carregue-o no Python utilizando bibliotecas como `pandas`.

Aplicação dos Algoritmos: Árvore de Decisão, KNN e SVM

Código para Árvore de Decisão

```
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
import pandas as pd
import matplotlib.pyplot as plt

# Carregar o dataset
df = pd.read_csv("seu_dataset.csv")
X = df[['coluna1', 'coluna2', 'coluna3']]  # Substitua pelas suas
    colunas
y = df['target']  # Coluna alvo

# Dividir os dados
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size
    =0.3, random_state=42)

#  rvore  de Decis o
clf_tree = DecisionTreeClassifier(max_depth=4, random_state=42)
clf_tree.fit(X_train, y_train)
y_pred_tree = clf_tree.predict(X_test)
print(f"Acur cia ( rvore  de Decis o): {accuracy_score(y_test,
    y_pred_tree):.2f}")

# Visualiza  o da  rvore
plt.figure(figsize=(16, 10))
plot_tree(clf_tree, feature_names=X.columns, class_names=['Classe1', '
    Classe2'], filled=True)
plt.show()
```

Código para K-Nearest Neighbors (KNN)

```
from sklearn.neighbors import KNeighborsClassifier

# KNN
clf_knn = KNeighborsClassifier(n_neighbors=5)
clf_knn.fit(X_train, y_train)
y_pred_knn = clf_knn.predict(X_test)
print(f"Acur cia (KNN): {accuracy_score(y_test, y_pred_knn):.2f}")
```

Código para Support Vector Machine (SVM)

```
from sklearn.svm import SVC

# SVM
clf_svm = SVC(kernel='linear', random_state=42)
```

```
clf_svm.fit(X_train, y_train)
y_pred_svm = clf_svm.predict(X_test)
print(f"Acur cia (SVM): {accuracy_score(y_test, y_pred_svm):.2f}")
```

Tarefa

1. Escolha um dataset no Kaggle ou UCI. 2. Aplique **Árvore de Decisão**, **KNN** e **SVM** para resolver o problema. 3. Compare os resultados dos três modelos em termos de acurácia. 4. Documente:

- O dataset escolhido (nome e fonte).
- O problema que está sendo resolvido.
- As métricas de desempenho para cada modelo.

Entrega

Entregue:

1. O código completo da árvore de decisão com 10 perguntas.
2. O código para a análise do dataset escolhido usando Árvore de Decisão, KNN e SVM.
3. Um relatório comparando o desempenho dos três modelos.

Boa sorte e divirta-se explorando algoritmos de aprendizado supervisionado!