

Support Vector Machines (SVM) - One-Versus-All (OVA)

Théorie

Les **machines à vecteurs de support (SVM)** sont des modèles de classification supervisés qui cherchent à maximiser la marge de séparation entre les classes. Pour un problème **multiclasse**, l'approche **One-Versus-All (OVA)** entraîne un SVM pour chaque classe, la comparant à toutes les autres classes combinées.

Hyperparamètre utilisé

Nous allons optimiser :

- **Paramètre de régularisation (C)** : contrôle la pénalisation des erreurs de classification et est sélectionné en fonction de la précision sur l'ensemble de validation.

Métriques d'évaluation

Nous afficherons :

- **Matrice de confusion** : montrant les erreurs de classification sur l'échantillon de test.
- **Taux de bien classés sur l'échantillon de validation** avec le meilleur hyperparamètre.
- **Taux de bien classés sur l'échantillon de test** avec ce même hyperparamètre.
- **Taux de bien classés par classe sur l'échantillon de test** pour observer la précision sur chaque classe.

Recherche du meilleur C et évaluation

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```

from sklearn.svm import SVC
from sklearn.multiclass import OneVsRestClassifier
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.preprocessing import StandardScaler
import warnings

# Suppression des avertissements inutiles
warnings.filterwarnings("ignore", category=UserWarning)

# Chargement des ensembles de données
train_data = pd.read_csv('../data/covertypes_train.csv')
val_data = pd.read_csv('../data/covertypes_val.csv')
test_data = pd.read_csv('../data/covertypes_test.csv')

# Préparation des données
X_train, y_train = train_data.drop('Cover_Type', axis=1),
    ↪ train_data['Cover_Type']
X_val, y_val = val_data.drop('Cover_Type', axis=1),
    ↪ val_data['Cover_Type']
X_test, y_test = test_data.drop('Cover_Type', axis=1),
    ↪ test_data['Cover_Type']

# Normalisation des données
scaler = StandardScaler()
X_train, X_val, X_test = scaler.fit_transform(X_train),
    ↪ scaler.transform(X_val), scaler.transform(X_test)

# Recherche du meilleur hyperparamètre C
C_values = np.arange(0.1, 1.1, 0.1)
val_accuracies = []

for C in C_values:
    model = OneVsRestClassifier(SVC(kernel='rbf', C=C))
    model.fit(X_train, y_train)
    acc = accuracy_score(y_val, model.predict(X_val))
    val_accuracies.append((C, acc))

# Sélection du meilleur hyperparamètre
best_C, best_val_acc = max(val_accuracies, key=lambda x: x[1])

# Affichage du graphique
plt.figure(figsize=(8, 6))
plt.plot(C_values, [acc for C, acc in val_accuracies],
    ↪ marker='o', linestyle='dashed', label="Validation")
plt.xlabel("Paramètre de régularisation (C)")

```

```

plt.ylabel("Précision sur validation")
plt.title("Impact de la régularisation sur la performance du SVM
↪ (OVA)")
plt.legend()
plt.show()

# Modèle final avec le meilleur hyperparamètre
final_model = OneVsRestClassifier(SVC(kernel='rbf', C=best_C))
final_model.fit(X_train, y_train)
y_test_pred = final_model.predict(X_test)

# Matrice de confusion
conf_matrix = confusion_matrix(y_test, y_test_pred)

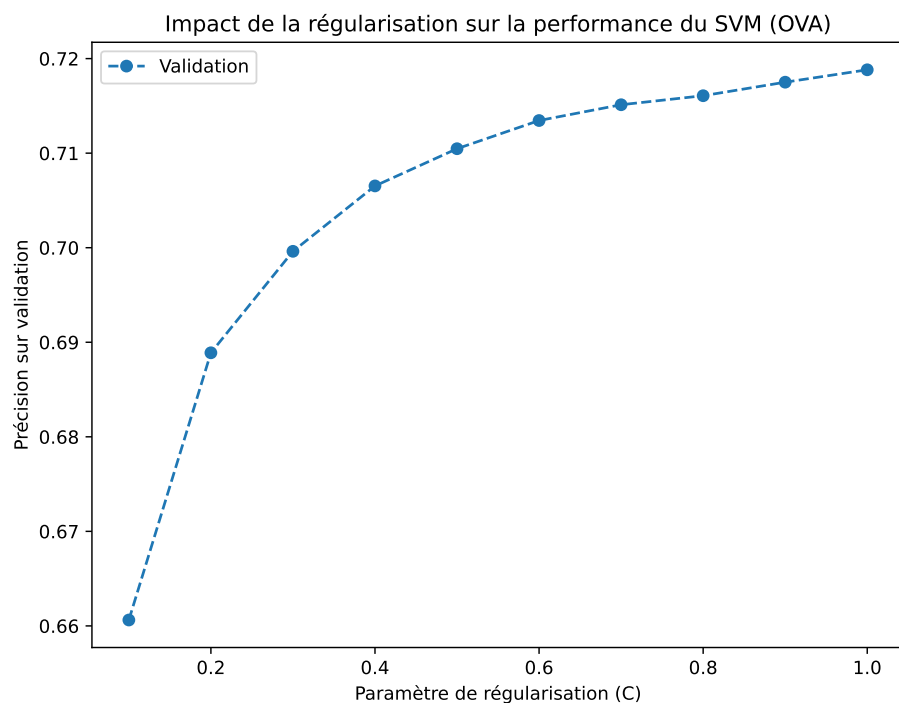
# Calcul des taux de bien classés par classe
class_accuracies = conf_matrix.diagonal() /
↪ conf_matrix.sum(axis=1)
overall_test_accuracy = accuracy_score(y_test, y_test_pred)

# Affichage des résultats
print(f"\n Meilleur hyperparamètre C sur l'échantillon de
↪ validation : {best_C:.2f}")
print(f"Taux de bien classés sur l'échantillon de validation avec
↪ cet hyperparamètre : {best_val_acc:.2%}")
print("\n Matrice de confusion sur l'échantillon de test, avec le
↪ meilleur hyperparamètre :")
print(conf_matrix)

print("\n Taux de bien classés par classe sur l'échantillon de
↪ test, avec le meilleur hyperparamètre :")
for i, acc in enumerate(class_accuracies, start=1):
    print(f"Classe {i} : {acc:.2%}")

print(f"\n Taux de bien classés sur l'échantillon de test avec le
↪ meilleur hyperparamètre : {overall_test_accuracy:.2%}")

```



Meilleur hyperparamètre C sur l'échantillon de validation : 1.00
Taux de bien classés sur l'échantillon de validation avec cet hyperparamètre : 71.88%

Matrice de confusion sur l'échantillon de test, avec le meilleur hyperparamètre :

```
[[1447  551    1    0    7    6  107]
 [ 424 2226   83    1   32   58    9]
 [    0   50 1296   18    4   62    0]
 [    0    0   81   22    0    7    0]
 [    5  202   30    0  126   17    0]
 [    0   47  378    3    0  266    0]
 [ 140    5    0    0    2    0  674]]
```

Taux de bien classés par classe sur l'échantillon de test, avec le meilleur hyperparamètre :

Classe 1 : 68.29%
Classe 2 : 78.57%
Classe 3 : 90.63%
Classe 4 : 20.00%
Classe 5 : 33.16%
Classe 6 : 38.33%
Classe 7 : 82.10%

Taux de bien classés sur l'échantillon de test avec le meilleur hyperparamètre : 72.22%