

TD SVM

Exercice 1

L'objectif de cette étude est de construire une meilleure séparation entre trois qualités des vins de bordeaux. Pour ce faire, un premier essai exécuté les commandes R suivantes :

```
> separateur = svm(quality ~., bordeaux)
> separateur

Call:
svm(formula = quality ~ ., data = bordeaux)
```

```
Parameters:
  SVM-Type:  C-classification
  SVM-Kernel: radial
    cost:    1
   gamma:   0.2
```

```
Number of Support Vectors: 32
```

1. Donner la signification des paramètres suivants : formula, data, SVM-TYPE, SVM-Kernel, cost, et Number of Support Vectors.

- Formula : Préciser la variable à prédire et les variables prédictives
- Data : Données d'apprentissage
- SVM-TYPE : Type de SVM (Classification ou régression)
- SVM-Kernel : Le type de la fonction noyau
- cost : Paramètre de régularisation
- Number of Support Vectors : Nombre des vecteurs supports

Le nombre de points supports est un bon indicateur. S'il est trop élevé par rapport à la taille de l'échantillon n , nous pouvons légitimement penser que la modélisation n'est pas très efficace.

2. En se basant sur les résultats suivants, calculer le taux d'erreur de classification du modèle obtenu :

```
> pred = predict(separateur, bordeaux)
> table(bordeaux$quality, pred)
      pred
      bad good medium
bad      11    0     1
good     0    11    0
medium  0     2     9
```

Taux de Bonne classification = 1 - Taux d'erreur = $((11+11+9) / (11+11+9+2+1)) * 100 = 91.17\%$

3. Spécifier la précision par qualité de vin.

Précision_Bad = $(11) / (11+0+0) * 100 = 100\%$

Précision_Good = $(11) / (11+0+2) * 100 = 84.61\%$

Précision_Medium = $(9) / (9+0+1) * 100 = 90\%$

4. Une deuxième étude a choisi de travailler avec une séparation polynomiale, et elle a donné les résultats suivants :

```
> separateur2 = svm(quality ~., bordeaux, kernel = "polynomial")
> pred = predict(separateur2, bordeaux)
> table(bordeaux$quality, pred)
      pred
      bad good medium
bad      10    0     2
good      0    5     6
medium    0    0    11
```

5. Calculer les taux de classification par qualité de vin, en déduire votre décision à propos le modèle le plus performant.

Taux de Bonne classification = $1 - \text{Taux d'erreur} = ((10+5+11) / (10+5+11+6+2)) * 100 = 76.47\%$

Précision_Bad = $(10) / (10+0+0) * 100 = 100\%$

Précision_Good = $(5) / (5+0+0) * 100 = 100\%$

Précision_Medium = $(11) / (11+6+2) * 100 = 57.89\%$