Université Batna 2 Master 2 SAD

Faculté de mathématiques et d’informatique 2020/2021

Département de mathématique

## Contrôle semestre 1

## Corrigé type

**Module** : Support vector machine

Questions de cours 5 pt

1. L'efficacité d'un SVM dépend de: D)

2. Les SVM sont moins efficaces lorsque: C) Lorsque les données ont du bruit et des points qui se chevauchent, il y a un problème pour dessiner un hyperplan clair sans mauvaise classification.)

3. Nous utilisons généralement la normalisation des fonctionnalités avant d'utiliser le noyau gaussien dans SVM. A,B

4. Supposons que vous ayez la même distribution de classes dans les données. Maintenant, disons que pour l'entraînement 1 fois en un contre tous les réglages, le SVM prend 10 secondes. Combien de secondes faudrait-il pour entraîner de bout en bout la méthode one-vs-all? B 10\*4=40

5. Supposons que votre problème ait changé maintenant. Désormais, les données n'ont que 2 classes. Que pensez-vous du nombre de fois que nous devons former SVM dans un tel cas? A

6. Un noyau en SVM est : A,B

7. La complexité temporelle minimale pour l’apprentissage d'un SVM est O(n2). D'après ce fait, quelles tailles d'ensembles de données ne conviennent pas le mieux aux SVM? A

8. C {\displaystyle C} 8.cC est une constante qui permet de contrôler le compromis entre nombre d'erreurs de classement, et la largeur de la marge. Plus elle est grande et plus cela revient à attribuer une forte importance à l’ajustement. Elle est le paramètre qui ajuste le compromis entre bon ajustement et bonne généralisation.

Elle doit être choisie par l'utilisateur, en général par une [recherche exhaustive](https://fr.wikipedia.org/wiki/Recherche_exhaustive) dans l'espace des paramètres, en utilisant par exemple la [validation croisée](https://fr.wikipedia.org/wiki/Validation_crois%C3%A9e) sur l'ensemble d'apprentissage. Le choix automatique de ce paramètre de régularisation est un problème statistique majeur.

Exercice 1 5pt

1. Maximiser l’équation $x^{2}+5xy+y^{2}$ sous la contrainte suivante : $x+2y=9$

$$L=f\left(x\right)-∝g\left(x\right)=x^{2}5xy+y^{2}-∝(x+2y-9)$$

$\frac{∂L}{∂x}=2x+5y-∝=0 \rightarrow ∝=2x+5y$ (1)

$\frac{∂L}{∂y}=5x+2y-2∝=0 \rightarrow ∝=\frac{5}{2}x+y$ (2)

$\frac{∂L}{∂∝}=x+2y-9=0$ (3)

De (1) et (2) $2x+5y=\frac{5}{2}x+y$ nous trouvons $x=8y$

Nous remplaçons dans (3) nous trouvons $y=\frac{9}{10}$ $x=\frac{36}{5}$

1. Maximiser l’équation $x^{2}+3xy+y^{2}$ sous la contrainte suivante : $x+y=8$

$$x^{2}+3xy+y^{2}-∝(x+y-8)$$

$\frac{∂L}{∂x}=2x+3y-∝=0 \rightarrow ∝=2x+3y$ (1)

$\frac{∂L}{∂y}=3x+2y-∝=0 \rightarrow ∝=3x+2y$ (2)

$\frac{∂L}{∂∝}= x+y-8=0$ (3)

$$x=y=4$$

Exercice 3 5pt

Soit l’ensemble des points de données suivant :

Points positives {(2,2) , (3,3) , (3,4) , (4,3)}

Points négatives {(1,1) , (0,-1) , (-1,1) , (-1,-1)}

S1=(1,1) S2=(2,2)

Calculer l’hyperplan optimal au sens de la SVM (w et b).

Tout d’abord augmenter les supports vecteurs

$$S\_{1}=\left(\genfrac{}{}{0pt}{}{1}{1}\right) \rightarrow \tilde{S}\_{1}=\left(\genfrac{}{}{0pt}{}{1}{\begin{array}{c}1\\1\end{array}}\right) $$

$$S\_{2}=\left(\genfrac{}{}{0pt}{}{2}{2}\right) \rightarrow \tilde{S}\_{2}=\left(\genfrac{}{}{0pt}{}{2}{\begin{array}{c}2\\1\end{array}}\right)$$

$$∝\_{1}S\_{1}S\_{1}+∝\_{2}S\_{1}S\_{2}=-1$$

$$∝\_{1}S\_{1}S\_{2}+∝\_{2}S\_{2}S\_{2}=1$$

$$3∝\_{1}+5∝\_{2}=-1$$

$$5∝\_{1}+9∝\_{2}=1$$

$$∝\_{1}=-7 ∝\_{2}=4$$

$$W=\sum\_{i=1}^{n}∝\_{i}S\_{i}$$

$$-7\left(\genfrac{}{}{0pt}{}{1}{\begin{array}{c}1\\1\end{array}}\right)+4\left(\genfrac{}{}{0pt}{}{2}{\begin{array}{c}2\\1\end{array}}\right)=\left(\genfrac{}{}{0pt}{}{1}{\begin{array}{c}1\\-3\end{array}}\right)$$

$$W=\left(\genfrac{}{}{0pt}{}{1}{1}\right), b=-3$$

Exercice 3 5pt

Corriger et commenter le code suivant :

#Import scikit-learn dataset library

from sklearn import data**sets**

#Load dataset

cancer == datasets.load\_breast\_cancer()

# Import train\_test\_split function

from sklearn.model\_selection import train\_test\_sp**lit**

# Split dataset into training set and test set

X\_train, X\_test, y\_train, y\_test = train\_test\_split(cancer.data, cancer.target, test\_size=0.3,random\_state=109) # 70% training and 30% test

#Import svm model

from sklearn import svm

#Create a svm Classifier with a linear kernel

clf = svm.SVC(kernel='linear')

#Train the model using the training sets

clf.fit(X\_train, **Y\_train**)

#Predict the response for test dataset

y\_pred = clf.predict(**X\_test**)

#Import scikit-learn metrics module for accuracy calculation

from sklearn import metrics

print("Accuracy:",metrics.accuracy\_score(y\_test, y\_pred))