SY19 - TD/TP 4

Automne 2013

Exercice 1

Soit l'ensemble d'apprentissage suivant dans $\mathbb{R}^2$: $\boldsymbol{x}_1 = (2,0)$ de la classe C_1 et $\boldsymbol{x}_2 = (0,2)$, $\boldsymbol{x}_3 = (-2,2)$, $\boldsymbol{x}_4 = (-1,3)$ de la classe C_2.

1. Donner l'équation (intuitivement) de *l'hyperplan* à vaste marge. Dessiner la frontière de décision correspondante. Vous indiquez quelles exemples sont des vecteurs support.
2. Calculer la marge optimale correspondante.
3. Déterminer la région de $\mathbb{R}^2$ pour laquelle une nouvelle observation venant dans l'ensemble d'apprentissage, et appartenant à C_1, est sans effet sur la solution (vous donnez l'équation et vous montrez sur la figure). Idem pour C_2.
4. Donner (en quelques lignes seulement) le code R qui permet d'aboutir à la solution que vous nommez "model".
5. Vous expliquez comment on peut confirmer les résultats intuitifs de la première question à partir du model obtenu (vous montrez les instructions qui permettent de donner les multiplicateurs de Lagrange et leurs valeurs numériques).

Exercice 2

On considère le problème des SVM vu en cours, dans le cas de données non linéairement séparable. A la place de la fonction $\sum_{i=1}^n \xi_i$ qui caractérise le coût des erreurs commises sur les éléments de l'ensemble d'apprentissage, on désire utiliser une autre fonction coût : $\frac{1}{2}\sum_{i=1}^n \xi_i^2$. Le problème d'optimisation consiste ainsi à minimiser

$$\frac{1}{2}\|\boldsymbol{w}\|^2 + \frac{c}{2}\sum_{i=1}^n \xi_i^2,$$

sous les contraintes :

$$\begin{aligned} y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i + w_0) &\geq 1 - \xi_i, \text{ pour } i = 1\cdots n \\ \xi_i &\geq 0, \text{ pour } i = 1\cdots n, \end{aligned}$$

où c étant un réel positif préalablement fixé.

1. Expliquer pourquoi les contraintes $\xi_i \geq 0$, pour $i = 1\cdots n$, sont inutiles.
2. En admettant que les contraintes $\xi_i \geq 0$, pour $i = 1\cdots n$, n'entrent pas dans l'équation du Lagrangien, construire le Lagrangien $L(\boldsymbol{w}, w_0, \boldsymbol{\alpha})$ où $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_n)$ constituent les multiplicateurs de Lagrange associés aux contraintes $y_i(\boldsymbol{w}\cdot\boldsymbol{x}_i - w_0) \geq 1 - \xi_i$, pour $i = 1\cdots n$.
3. Calculer $\frac{\partial L}{\partial \boldsymbol{w}}$, $\frac{\partial L}{\partial w_0}$ et $\frac{\partial L}{\partial \xi_i}$, et calculer au point selle, $\boldsymbol{w}$ et ξ_i en fonction de y_i, α_i et $\boldsymbol{x}_i$.
4. Trouver ainsi le Lagrangien dual $W(\boldsymbol{\alpha})$, avec les contraintes correspondantes et montrer que le problème peut être ramené à un problème d'optimisation quadratique sous contraintes de la forme $W(\boldsymbol{\alpha}) = \sum_{i=1}^n \alpha_i - \frac{1}{2}\sum_{i,j=1}^n \alpha_i\,\alpha_j[\,y_i\,y_j\;(\boldsymbol{x}_i\cdot\boldsymbol{x}_j) + \frac{\delta_{ij}}{c}]$ où $\delta_{ij} = 1$ si $i = j$ et $\delta_{ij} = 0$ si $i \neq j$.

5. Ecrire les conditions de Kuhn-Tücker qu'on doit satisfaire à l'optimalité et déduire que tous les vecteurs de support (qui ont des $\alpha_i > 0$) auront nécessairement un $\xi_i > 0$.
6. Peut on généraliser cette méthode pour obtenir des frontières de décision non-linéaires dans l'espace d'origine ?. Pourquoi et comment ? Donner alors la forme finale de la fonction de décision.

Exercice 3

Soit l'ensemble d'apprentissage suivant dans $\mathbb{R}$ où $\boldsymbol{x}_1 = 1$, $\boldsymbol{x}_2 = 2$, $\boldsymbol{x}_3 = 6$ de la classe C_1 ($y_i = 1$, $i \in \{1,2,3\}$) et $\boldsymbol{x}_4 = 4$, $\boldsymbol{x}_5 = 5$ de la classe C_2 ($y_i = -1$, $i \in \{4,5\}$). On considère le problème de classification par support vector machines étudié dans le cours, où $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{\phi}(\boldsymbol{x}_i).\boldsymbol{\phi}(\boldsymbol{x}_j) = (\boldsymbol{x}_i.\boldsymbol{x}_j + 1)^2$ et $c = 100$.

1. Donner l'expression du Lagrangien à maximiser (avec les contraintes associées).
2. L'utilisation d'un programme de résolution de problème quadratique a permis l'obtention des α_i, $i \in 1..5$. Parmi l'ensemble de propositions ci après, choisir la solution possible. Vous devez commenter les propositions rejetées.
   (a) $\alpha_1 = 2, \alpha_2 = -1, \alpha_3 = 0, \alpha_4 = 1, \alpha_5 = 1$;
   (b) $\alpha_1 = 0, \alpha_2 = 3, \alpha_3 = 1, \alpha_4 = 2, \alpha_5 = 1$;
   (c) $\alpha_1 = 0, \alpha_2 = 2.5, \alpha_3 = 4.833, \alpha_4 = 0, \alpha_5 = 7.333$;
   (d) $\alpha_1 = 101, \alpha_2 = 2, \alpha_3 = 0, \alpha_4 = 102, \alpha_5 = 1$;
3. Donner (en quelques lignes seulement) le code R qui permet d'aboutir à la solution choisie.
4. Donnez l'expression de la fonction discriminante $g(\boldsymbol{x}) = \boldsymbol{w}^* \cdot \boldsymbol{\phi}(\boldsymbol{x}) + w_0^*$ et donner la règle de décision résultante.
5. Dessiner $g(\boldsymbol{x})$ en fonction de $\boldsymbol{x}$ ($\boldsymbol{x}$ est dans $\mathbb{R}$ même s'il est écrit en gras). Noter les points de l'ensemble d'apprentissage sur la figure en précisant clairement les "vecteurs support" et en représentant sur le dessin votre règle de décision.

Exercice 4

Objectif : Mettre en œuvre la méthode des séparateurs à vaste marge, en testant l'influence du choix des paramètres.
On désire appliquer la méthode sur les données du fichier breast-cancer-wisconsin.data dans le but discriminer les tumeurs bénignes des tumeurs malignes en utilisant des données morphologiques. L'ensemble d'exemples disponible est constitué de 699 cas correctement classifiés.

Travail Préliminaire : Manipulation et compréhension des données

Avant de commencer, il vaut mieux ouvrir le fichier *breast-cancer-wisconsin.data* dans un éditeur de texte, de préférence *wordpad*, et regarder les données.
Ci-après quelques opérations et commandes en R, avec les commentaires explicatifs, pour vous aider à manipuler et comprendre les données du fichier *breast-cancer-wisconsin.data*.

```
####################################################

#lire " breast cancer data " :

bcdata <- read.csv('breast-cancer-wisconsin.data',head=TRUE)

#Regarder les noms des colonnes.

names(bcdata)

#regarder une colonne, par exemple:
```

```
bcdata$ClumpThickness #ou bcdata$Samplecodenumber

#regarder les classes:

bcdata$Class

#Enlever de bcdata la colonne "Samplecodenumber" et la colonne
#"Class". Le signe moins pour dire unselect les deux colonnes:

databcall <- subset(bcdata,select=c(-Samplecodenumber,-Class))

# vous pouvez faire names(databcall)

# Selectionner les classes seulement:

classesbcall <- subset(bcdata,select=Class)

#prendre une partie de databcall pour l'apprentissage:

databctrain <- databcall[1:400,]

classesbctrain <- classesbcall[1:400,]

#prendre une partie de databcall pour le test:

databctest <- databcall[401:699,]

classesbctest <- classesbcall[401:699,]


##################################################
```

Application des SVM sous R

La fonction du langage R à utiliser est *svm* du package e1071. Un exemple d'application aux données Breast-Cancer est le suivant :

```
##################################################


# Faire help(svm) pour bien comprendre la fonction "svm" :

model <- svm(databctrain, classesbctrain)

# Faire str(model)

# Faire help(predict.svm) #Attention, ici "model" est un objet de
#la classe "svm". Ainsi, il faut faire help(predict.svm) et non pas
#help(predict)

# Validation de "model" :

pred <- predict(model, databctest)
```

```
# Comparer la prediction et les vraies classes :

table(pred,t(classesbctest))

#Les hyperparametres affecte les performances du noyau. Le
#package e1071 offre une fonction, tune(), qui fait ce qu'on
#appelle "grid search" et donne ainsi une estimation des
#parametres. Pour cela, faire : help(tune)

#Exemple d'application de la fonction "tune" :

a=tune(svm, train.x=databctrain, train.y=classesbctrain,
validation.x=databctest, validation.y=classesbctest, ranges =
list(gamma = 10^(-1:1), cost = c(1,1.5,2)), control =
tune.control(sampling = "fix"))

# Faire  str(a)

model <- svm(databctrain,
classesbctrain,gamma=a$best.parameters$gamma,cost=a$best.parameters$cost)


# Comparaison de la prediction et des vraies classes

pred <- predict(model, databctest)

table(pred,t(classesbctest))


##################################################
```

Travail demandé

La mise en œuvre des SVM nécessite un certain nombre de choix et de réglages. La fonction *tune* présentée ci-dessus permet de faire ce réglage pour un noyau choisi a priori. Il est cependant important que vous testiez l'influence de ces différents choix sur la qualité de vos résultats. Par la suite, vous utilisez l'erreur moyenne comme critère de mesure de performance.

1. Montrer la différence entre les expressions des noyaux polynomial et gaussien de la fonction SVM et celles vues en cours.
2. Pour un noyau gaussien avec une largeur de bande raisonnablement choisie, montrer sur une figure la variation de la probabilité d'erreur en fonction du paramètre de pénalisation γ ; Commenter le résultat. Faire une comparaison entre les "boxplot" obtenus sur les composantes du vecteur $\boldsymbol{\alpha}$ (pour différentes valeurs de γ) et commenter le résultat. Montrer sur une figure la variation du nombre des exemples qui ont des α_i non nuls en fonction de γ.
3. Dans cette partie on met $\gamma = \infty$. Montrer sur une figure l'influence du choix du paramètre associé au noyau sur les performances de la méthode(ici c'est le choix de la *largeur de bande* du noyau gaussien). Montrer aussi, en fonction de ce choix, la variation du nombre des exemples qui ont des α_i non nuls. Commenter les différents résultats.
4. En se servant de la fonction "tune", et en fixant $\gamma = 1$, comparer les résultats obtenus par un noyau polynomial et un noyau gaussien. Commenter les résultats.

Exercice 5 : Travail Supplémentaire

Pour les plus motivés, vous pouvez utiliser la fonction *ksvm* du package *kernlab* pour répondre au problème de reconnaissance de caractères. Vous pouvez aussi utiliser d'autres fonctions de ce même package telles que la fonction *kpca*, pour l'analyse en composante principale à noyau.