



Aide mémoire R

Quelques librairies R

Liens &amp; contact

Recherche : 

|                            |
|----------------------------|
| Courbe ROC                 |
| Arbres de décision (rpart) |
| Arbres de décision (tree)  |
| Random Forest              |
| Réseaux de neurones        |
| SVM                        |

Aide-memoire-R &gt; Apprentissage &gt; SVM

## SVM.

SVM : Support Vector Machines. Ce sont des technologies qui permettent un apprentissage en limitant au maximum les risques d'over-fitting (sur des bases théoriques). On utilisera ici la librairie `e1071` (faire `library(e1071)`, la librairie devant être installée avant) qui s'appuie sur `libsvm`.

Phase d'apprentissage :

- soit avec une forme traditionnelle, en fournissant les variables prédictives, puis la variable à prédire : `model <- svm(subset(iris, select = c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width")), iris$Species)` (cf le data frame "iris" qui peut être obtenu en faisant `data(iris)`).
- soit avec une formule : `model <- svm(Species ~ ., iris)` (le '.' indique que toutes les variables autres que Species doivent être utilisées).

Par défaut, le SVM utilisé est de type à noyau radial et les valeurs sont normalisées avant de faire tourner le SVM (pour avoir une moyenne 0 et un écart-type 1). L'objet retourné est de classe `svm`.

Paramètres du SVM :

```
model <- svm(Species ~ ., iris, type = "C-classification", kernel = "polynomial", degree = 2, gamma = 2, cost = 0.5, coef0 = 0)
```

- type : C-classification si on fait de la classification simple (défaut si la variable prédite est de type factor), epsilon-regression si on fait de la régression (défaut si la variable prédite est de type numérique).
- kernel : noyau à utiliser : radial (radial basis de type  $\exp(-\gamma \|u-v\|^2)$ ), linear ( $u \cdot v$ ), polynomial:  $(\gamma u \cdot v + \text{coef0})^{\text{degree}}$ , sigmoid:  $\tanh(\gamma u \cdot v + \text{coef0})$ .
- degree : degré du polynôme (pour kernel polynomial).
- gamma : facteur gamma (inutile pour kernel linéaire).
- coef0 : coefficient pour kernel polynomial ou sigmoïde.
- cost : constante de coût pour le lagrangien.

Attributs d'un objet de la classe `svm` :

- `model$SV` : matrice des vecteurs de support.
- `model$nSV` : nombre de vecteurs de support.

Par défaut, les valeurs prédites sont calculées et incluses dans le modèle. On peut alors comparer les valeurs prédites et les valeurs réelles :

```
model <- svm(Species ~ ., iris, fitted = TRUE)
table(model$fitted, iris$Species)
      setosa versicolor virginica
setosa      50           0         0
versicolor   0          48         2
virginica    0           2        48
```

Cross-validation :

- Faire `model <- svm(Species ~ ., iris, cross = 10)` pour faire des cross-validation sur 10 fois.
- On peut alors récupérer le vecteur des précisions de prédictions obtenues par `model$accuracies`.

Prédiction de classes :

- `predict(model, subset(iris, select = - c(Species)))` : fait la prédiction sur le dataframe ou la matrice donnée et renvoie un vecteur de prédictions.
- `table(predict(model, subset(iris, select = - c(Species))), iris$Species)` : compare les valeurs prédites et les valeurs réelles.

Dernière date de modification : 2013-12-07