

BI - TP1 - Classification et protocole expérimental en RapidMiner

Ludovic Denoyer

1 Protocole expérimental classique

Données : <https://www.kaggle.com/c/titanic/data> – utiliser train.csv

Nous allons nous intéresser à faire des arbres de décision sur le dataset **titanic** afin de prédire si un passager survit ou non. Le protocole expérimental consiste à découper le dataset en train et test, et à évaluer (sur le train et test) l'accuracy du modèle. **Outils** : Split Data, Decision Tree, Apply Model, Performance (classification)

1.1 Protocole expérimental (v2)

- Utiliser l'outil "Split Validation" pour effectuer le même travail
- Utiliser l'outil "X-Validation" pour effectuer le même travail

1.2 Optimisation d'hyper paramètres

Dans "X-Validation", utiliser l'outil "Optimize Parameters" afin d'optimiser la profondeur maximale de l'arbre de décision

1.3 Question....

Quels sont les critères clefs permettant de prédire la survie d'un passager du titanic ?

2 Modèles continus

Nous allons effectuer le même travail sur les données "housing" pour la tâche de régression. Nous allons tester plusieurs méthodes de régression afin de déterminer les facteurs les plus importants permettant de déterminer le prix des maisons. Les données sont disponibles ici : <http://mldata.org/repository/data/viewslug/regression-datasets-housing/>

3 Sélection de variables

3.1 Sélection greedy

- Utiliser l’outil ”forward selection” ainsi que les ”decision trees” pour sélectionner les variables les plus pertinentes du dataset titanic
- Faire la même chose à l’aide d’un SVM en rendant continues les variables qualitatives.

3.2 PCA

Sur MNIST (<http://pjreddie.com/projects/mnist-in-csv/>) – en utilisant le test uniquement et en le splittant en deux :

- Tester la performance d’un SVM
- Utilisez une PCA pour réduire le nombre de dimensions (à 2). Visualiser le nouveau dataset obtenu
- Utilisez un k-NN pour calculer la performance de classification
- Testez les cartes SOM sur le même contexte
- Utilisez l’outil ”Visualize model by SOM”
- Qu’en déduisez vous ?