

Survival and longitudinal data analysis

Exercice 1

Survival analysis or classification ?

Le jeu de données **wpbc** (disponible sur <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wpbc.data>). Il est présenté sur <https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/wpbc.names>.

On souhaite prévoir la probabilité de rechute (“recurrent”) à 24 mois. Pour cela, vous comparerez les méthodes de l’analyse de survie (modèles de Cox, survival random forests, ...) aux méthodes de classification. Les mesures de performances (notamment l’AUC) se feront sur un sous-échantillon de test formé de 20 à 30% des données (attention à bien stratifier !).