

## SY19 – A16

### TP 4 (noté) : Régression et classification – Sélection de modèles

Le but de ce TP est de traiter :

- Un problème de classification : reconnaissance de phonèmes (données **Phonemes**) ;
- Un problème de régression : prédiction du temps de récurrence du cancer du sein (données **Breast cancer**).

Dans les deux cas, il faudra construire des prédicteurs ayant les meilleures performances possibles, et les évaluer.

Vous êtes libres de la méthodologie à employer. Les critères de notation seront :

1. Variété des méthodes utilisées, correction et rigueur dans l'utilisation des méthodes ;
2. Profondeur et pertinence de l'analyse des résultats ;
3. Qualité du rendu écrit (clarté des explications ; correction du français ou de l'anglais ; qualité des tableaux et des figures ; soin dans la présentation du rapport).

Documents à rendre au plus tard le 22 novembre à minuit sur Moodle :

1. Rapport écrit, en français ou en anglais (de préférence en Latex) ;
2. Code source R commenté.

Toute similitude non fortuite entre deux rapports se traduira par la multiplication des notes par un facteur compris entre 0 et 1/2.