

## Projet final

Ce devoir est à réaliser par groupe de 1, 2, ou 3 élèves. Il faudra rendre un rapport en pdf, un fichier texte pour les prédictions du premier exercice et les codes utilisés en annexe.

### 1 Régression binaire

Le fichier “simu.txt” contient une simulation d’un modèle binaire :

- 2000 observations.
- 2 variables explicatives  $X_1$  et  $X_2$ .
- La variable à expliquer  $Y$  est à valeurs dans  $\{1, 2\}$ .

1. Déterminer le meilleur modèle possible pour modéliser  $P(Y|X_1, X_2)$ .
2. Utiliser le fichier “xsimutest” pour prédire 1000 observations de la variable  $Y$  du fichier de test. Il faut sauvegarder ces prédictions dans un fichier “.txt”, les séparateurs devront être des retours chariot. Par exemple, si les 1000 prédictions du fichier de test de  $Y$  sont dans le vecteur “predictions”, on le sauvera avec la procédure R :

```
write.table(predictions,"predictions.txt",row.names=F,col.names=F)
```

Il faudra joindre ce fichier à votre devoir.

### 2 Races de chiens

On dispose du fichier “chiens” qui détaille les caractéristiques des chiens en fonction de la race. Les variables sont les suivantes :

- TAI (Taille) : 3 = *grande*, 2 = *moyenne* et 1 = *petite*.
- POI (Poids) : 3 = *lourd*, 2 = *moyen* et 1 = *lger*.
- VEL (Vélocité) : 3 = *rapide*, 2 = *moyenne* et 1 = *lent*.
- INT (Intelligence) : 3 = *grande*, 2 = *moyenne* et 1 = *faible*.
- AFF (Affectuosité) : 2 = *grande* et 1 = *faible*.
- AGR (Agressivité) : 2 = *grande* et 1 = *faible*.
- FON (Fonction) : 3 = *garde*, 2 = *chasse* et 1 = *compagnie*.

1. En prenant la variable FON comme variable supplémentaire, faire une anayse des correspondances multiples de ces données.
2. En déduire une description des différentes races de chiens.