

## TD 2

1. On utilise les données Weekly du package ISLR. Cet ensemble de données est similaire à celui de Smarket, avec la différence qu'il contient 1089 rendements hebdomadaires sur 21 ans, de début 1990 à fin 2010.
  - a) Produire quelques résumés numériques et graphiques des données hebdomadaires. Quelles caractéristiques des données détectez-vous?
  - b) Utiliser l'ensemble des données pour estimer un modèle de régression logistique avec Direction comme variable à expliquer et les cinq rendements retardés plus Volume comme prédicteurs. Utiliser la fonction summary pour afficher les résultats. Quels prédicteurs expliquent significativement la direction du marché ?
  - c) Calculer la matrice de confusion et la fraction des bonnes classifications. Expliquer les types d'erreur renseignés par la matrice de confusion.
  - d) Maintenant, estimer le modèle de régression logistique sur un ensemble d'apprentissage sur la période allant de 1990 à 2008, en utilisant Lag2 comme unique prédicteur. Calculer la matrice de confusion et la fraction des bonnes classifications sur l'échantillon test (càd les données de 2009 à 2010).
  - e) Refaire d) avec la LDA
  - f) Refaire d) avec la QDA
  - g) Refaire d) avec KNN en utilisant  $K = 1$
  - h) Quelle est la méthode qui donne les meilleurs résultats sur cet ensemble de données ?
  - i) Refaire d) avec KNN en utilisant différentes valeurs de  $K$ .