elasti linias as taici. - liglas brendi fidale frama Later - Pales

Université Echahid Hama Lakhdar, El-oued Institut des sciences exactes Departement d'informatique

 $1^{er}Master$: Intelligence Artificielle et Systémes distribués

Semestre: 1.2022 Mohammed Anguar Nagui

— TP : Analyse de données —

1 Exercice1: 12pts

On considère les données, data.csv, train.csv et test.csv tell que :

- data.csv : Les données totales (Figure 1).
- train.csv : Les données utilisées pour la création du modèle.
- test.csv : les données utilisées pour tester le modèle.

Les attribues des données X1 et X2 numérique et $y \in \{0,1\}$

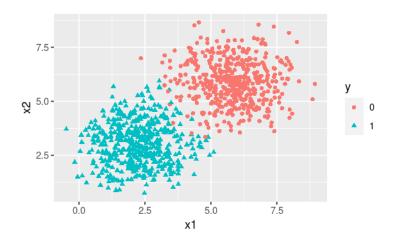


Figure 1 –

- 1. Construire le modèle (linear model) de régression linéaire multiple de la forme y = f(X1, X2).
- 2. L'hypothèse paramétrique gaussienne de la forme de régression d'analyse quadratique discriminante $X/Y = k \sim \mathcal{N}(\mu_k, \Sigma_k)$

$$\mathbb{P}(x/Y = k) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{\frac{1}{2}}} exp(-\frac{1}{2}(x - \mu_k)^T \Sigma_k^{-1}(x - \mu_k))$$
 (1)

Où p est le nombre de variables explicative. Construire le modèle d'analyse quadratique discriminante (model qda).

- 3. En analyse discriminante linéaire, les matrices de covariance sont supposées égales. Estimer sur les données d'apprentissage la matrice $\Sigma = \Sigma_1 = \Sigma_2$. Construire le modèle d'analyse discriminante lineaire (model lda)
- 4. Prédire la classe de X=(2.75,1.5)
 - Par le modèle linear model.
 - Par le modèle **model qda**.
 - Par le modèle **model** lda.
- 5. Prédire les classes de données test.csv
 - Par le modèle linear model.
 - Par le modèle model_qda.
 - Par le modèle **model lda**.
- 6. On veut évaluer les résultats des modèles **linear_model**, **model_qda** et **model_lda** par deux méthodes Confusion Matrix (Confusion Matrix) et la courbe CRO (ROC curve).
 - Évaluer la qualité des modèles linear_model model_qda et model_lda par Confusion Matrix et ROC curve.

- 7. Quel est le modèle le plus judicieux {argumenter votre réponse}.
- 8. Que-est-ce-que la Frontière de décision (decision boundaries).
- 9. Dessiner la Frontière de décision par la fonction **decisionplot** (dans le fichier function_decision_boundaries)pour le modèle **model qda** :

et pour le modèle model lda:

- 10. Dessiner la Frontière de décision de modèle model_linear s'il exist et comparer avec les Frontières de décisions de model_qda et model_lda.
- 11. Après la fin de ce Tp comment voyer vous le modèle linéaire et model qda et model lda

2 Exercie2: 4pts

Dans un problème de régression linéare simple on optimise le carré des sommes des erreuers résiduelles $e_i, Min(\sum_{i=1}^{i=n}(e_i)^2)$, sachant que : $b=\bar{y}-a\bar{x}$. Puisuqe $e_{(\bar{x},\bar{y})}=0$.

Ecrire le code R qui permet d'estimer a et b., Tell que :

$$\begin{cases} y_i = ax_i + b + e_i \\ Min(\sum_{i=1}^{i=n} (e_i)^2) \\ b = \bar{y} - a\bar{x} - e_{(\bar{x}, \bar{y})} \\ e_{(\bar{x}, \bar{y})} \neq 0 \end{cases}$$

-Comparer avec le modèle de régression lineare simple vue en cours.

3 La qualité de présentation des réponses 04 pts

- Pour assurer la meilleure présentation, je vous conseille d'utiliser : R notebook et LATEX.

Data set et la fonction de la Frontière de décision

Les fichiers, data.csv, train.csv, test.csv et function_decision_boundaries. https://github.com/manouarn/Data-Analysis-home-work-2023