

Rapport

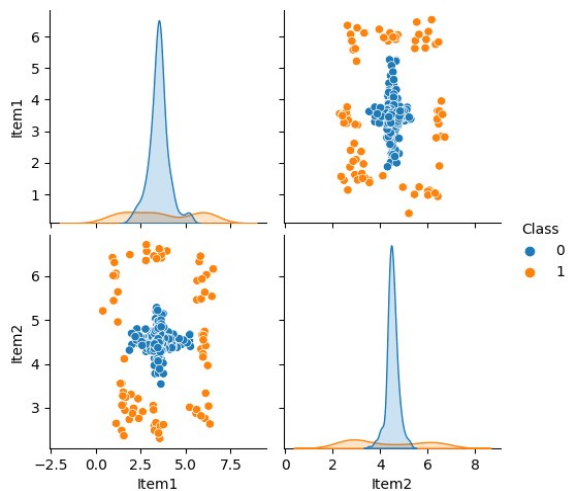
I- Données

1-Nombre de données dans chaque classe:

-Inlier: 250 données

-Outliers: 80 données

2-Capture d'écran du résultat:



3-la forme des données: les données nous permettent pas de dresser une fonction pour distinguer ou si j'ose dire classifier les données.

II- Évaluation

1. À la seule lecture des coefficients, le modèle me semble bon car le modèle a prédit 1000 valeurs qui sont vraies négatives qui sont des inliers, ces derniers sont largement supérieurs à la somme des valeurs restantes 2+30+5, mais je peux aussi rajouter que le modèle n'est pas bon car le nombre de vraies positives (outliers prédit comme outliers) est fortement inférieurs aux valeurs fausses négatives (outliers prédits comme inliers) donc le modèle ne prédit pas bien les outliers car il a classé 30 valeurs sur 35 en tant que inliers.

2- Calcule comparaison de l'exactitude et l'exactitude pondérée de la matrice confusion:

$$\text{exactitude} = (T N + T P) / (T N + F P + F N + T P) = (1000 + 5) /$$

$$(1000+2+30+5)=0.969=96,9\%$$

$$\text{exactitude pondérée} = ((TN)/(TN+FP) + (TP)/(FN+TP))/2 = ((1000)/(1000+2) + (5)/(30+5))/2 = 0.57 = 57\%$$

L'exactitude est 96,9 qui est un bon score qui rapproche des 100% et largement supérieur de l'exactitude pondérée qui est de 56%

3- L'exactitude donne un score aussi bon car le modèle prédit bien les inliers qui sont largement supérieur en nombre par rapport aux outliers qui sont que 35 outliers dans le modèle ,malgré le modèle ne prédit pas bien les outliers mais tellement le modèle prédit bien les inliers et aussi y a une grande différence dans le nombre des inliers par rapport aux outliers 1002 inliers > 35 outliers donc l'exactitude est bon

4-Elle n'est pas pertinente dans notre cas car nous on chercher à bien prédire les outliers et les différenciers des inliers, mais le modèle fait le contraire il prédit 30 outliers comme inliers mais que 5 outliers comme outliers.cela dit le modèle n'est pas pertinent dans notre cas.