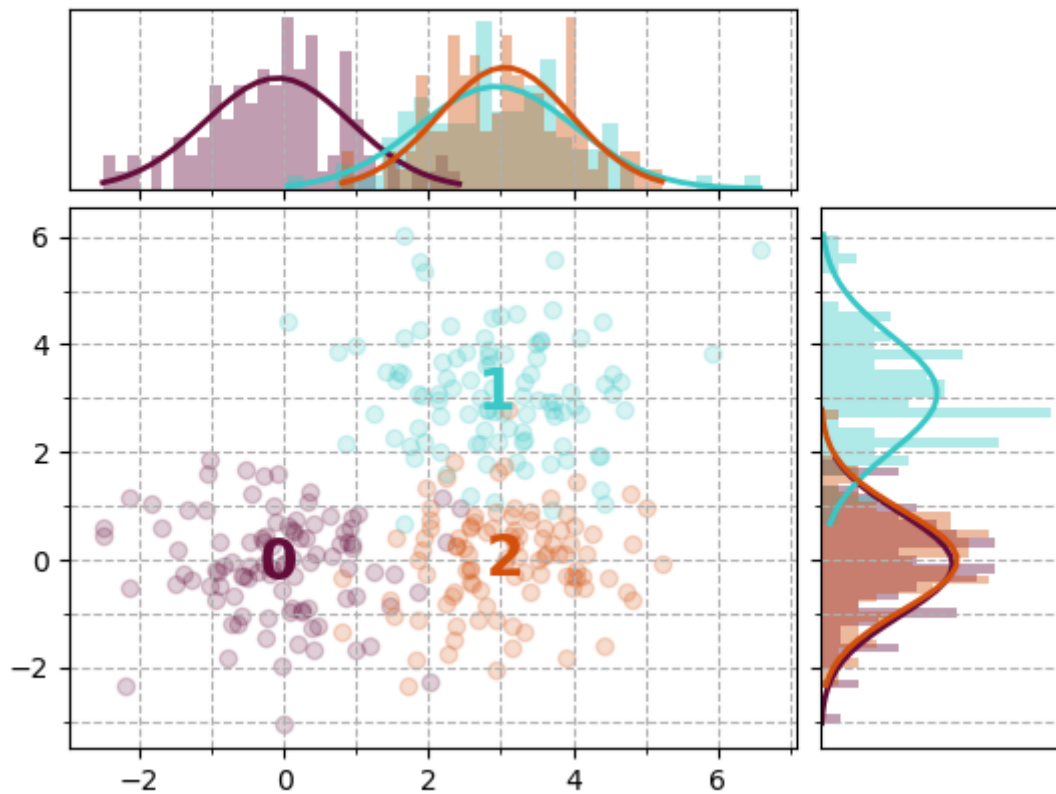


TP3 : Classification par approche Bayésienne

1 Analyse de données



2 Apprentissage des paramètres de la loi normale

Dans la fonction fit(), on va calculer le vecteur moyen ainsi que la matrice de covariance. Nous pouvons utiliser des fonctions de numpy pour ce faire.

Pour le vecteur moyen :

```
for i in range(n_classes):
    #calcul vecteur moyen
    self.mu[i] = np.average(X[y == i], 0)
```

np.average permet de faire une moyenne.

Pour la matrice :

```
for i in range(n_classes):
    #calcul de la matrice
    self.sigma[i] = np.cov(X[y == i].T)
```

np.cov permet d'obtenir la covariance.

3 Classification : Décision par maximum de vraisemblance

```
#calcul des determinants
for k in range(n_classes):
    det[k] = (np.linalg.det(self.sigma[k]))

for i in range(n_obs):
    maxVal = -100000
    print("Pour le point", i)
    for j in range(n_classes):
        print("Pour la classe", j)
        val = -(0.5 * math.log(det[j])) - (0.5)*np.dot(np.dot(np.transpose(X[i] - self.mu[j]), (np.linalg.inv(self.sigma[j])), X[i] - self.mu[j]))
        print(val)

    #mise à jour classe
    if(maxVal < (val)):
        maxVal = (val)
        y[i] = j
```

On utilise la forme transformée de la formule avec le log. Dans un soucis d'optimisation les déterminants sont calculés une seule fois. On utilise les fonction numpy comme linalg.det pour obtenir le déterminant ou encore transpose pour la transposée.

Avec cette formule les résultats sont satisfaisant car on obtient pour la classification des trois points du fichier test une précision de 100%

```
predict [ 0.  1.  2.]
precision : 1.00
```

4 Probabilité à priori

La probabilité a priori se définit au moment de l'instanciation de la classe GaussianB.

```
# Instanciation de la classe GaussianB
g = GaussianBayes(priors=[0.00001,0.00001,0.8])
```

Après calcul de la formule, on ajoute la probabilité à priori.

```
#ajout de la probabilité à priori
val = val * self.priors[j]
```

5 Généralisation

Le programme peut réaliser la classification avec un nombre quelconque de classes car il n'y a aucune instruction qui limite le nombre de classes. La même chose pour la dimension des données, elle peut être variable.