

## QUIZZ 1

## Exercice 1 – IRIS

Un des problèmes les plus classiques en ML est IRIS, une base de classification des fleurs iris en 3 classes (sous-familles des iris) à partir de 4 caractéristiques (*Sepal Length*, *Sepal Width*, *Petal Length* and *Petal Width*).

Les données sont déjà chargées dans l'environnement et stockées classiquement dans une matrice  $X \in \mathbb{R}^{N \times 4}$  et les étiquettes dans un vecteur  $Y \in \{0, 1, 2\}^N$

**Q 1.1 (1pt)** En choisissant une modélisation bayésienne naïve basée sur la loi normale, combien de paramètres faut-il apprendre ?

**Q 1.2 (3pts)** Donner le code permettant de tracer les 3 figures de type *scatter plot* correspondant à la variable 1 par rapport aux trois autres ( $X_1$  vs  $X_2$ ,  $X_1$  vs  $X_3$ ,  $X_1$  vs  $X_4$ ) en coloriant les points en fonction de leur classe d'appartenance.

**Objectif :** extraire les bonnes lignes et colonnes à partir des matrices **X,Y** de la question précédente.

```
1 figure()
2 scatter(x, y) # x: vecteur contenant les abscisses; y: vecteur des ordonnées
3               # plusieurs appels dans la même figure => différentes couleurs
4 # ou
5 scatter(x,y,c) # c : couleur des points = soit 1 valeur;
6               # soit un vecteur contenant une valeur par point
```

**Q 1.3 (2pt)** Etant donné un modèle qui a été entraîné et appliqué sur des données, nous avons un vecteur  $Y_{gt} \in \mathbb{R}^N$  contenant la vérité terrain et  $Y_{hat} \in \mathbb{R}^{N \times 3}$  contenant nos prédictions en probabilité au sens du maximum de vraisemblance  $Y_{hat} = P(X|Classe)$ . Donner la/les ligne(s) de code permettant de calculer le taux de bonne classification à partir de  $Y_{gt}$  et  $Y_{hat}$ . **Objectif :** travail sur `np.where` et la recherche de maximum