1. Le fichier `XYtrain.dat` contient des données de classification binaire du type

```
-0.8633 -2.2381 1
-1.4543 -3.1064 1
...
2.5098 1.8009 0
2.2547 -0.0484 0
```

où les deux premières colonnes sont les coordonnées de points dans $\mathbb{R}^2$, et la dernière colonne représente la classe (0 ou 1).

   1. Représenter visuellement ces données, en coloriant les points en fonction de l'appartenance à l'une des classes
   2. Construire un classifieur kNN avec l'ensemble ce des données, en imposant $k = 7$. Mesurer l'erreur de test et l'erreur d'entraînement résultantes.
   3. Au lieu d'utiliser toutes les données, choisir aléatoirement $n = 5$ données dans chaque classe, et entraîner le classifieur kNN avec ces 10 données, toujours avec $k = 7$. Mesurer l'erreur de test et l'erreur d'entraînement résultantes. Comparer avec l'erreur obtenue en utilisant toutes les données et commenter le résultat.
   4. En rajoutant progressivement de nouvelles données aléatoires, faire varier n de 10 à 50 par incrément de 5. A chaque étape, entraîner le classifieur kNN avec ces données, toujours avec $k = 7$. Tracer ensuite les courbes d'erreur (test et entraînement) en fonction de n
   5. Au lieu de rajouter les nouveaux points aléatoirement, utiliser une stratégie d'apprentissage actif pour faire varier n de 10 à 50 par incrément de 5. Tracer ensuite les courbes d'erreur (test et entraînement) en fonction de n et comparer avec les résultats obtenus précédemment. L'apprentissage actif est-il utile ?

2. 1. Dans $\mathbb{R}^2$, générer deux groupes de 500 points distribués selon des lois normales centrées respectivement en $(1, 0)$ et en $(0, 1)$, avec un écart-type $\sigma = 0.1$, et prendre 150 points aléatoirement dans chaque groupe comme ensemble de test.
   2. Utiliser un SVM linéaire dans le cadre de l'apprentissage "passif" en ajoutant petit à petit (e.g., par groupe de 10) les 350 points dans l'ensemble d'entraînement, et représenter l'évolution des erreurs d'entraînement et de test en fonction du nombre de points utilisés pour l'apprentissage.
   3. Répéter le calcul avec une stratégie d'apprentissage actif, et comparer les courbes d'erreur.
   4. Répéter ces opérations avec $\sigma = 0.2$, $\sigma = 0.4$, $\sigma = 0.6$ et comparer les résultats
   5. Refaire cette comparaison dans $\mathbb{R}^n$, avec $n = 3, 4, 5, 10, 20, 30, 50, 100$. Les jeux de données sont également des normales, centrées aux extrémités des vecteurs unitaires sur les axes de $\mathbb{R}^n$. Commenter le résultat.