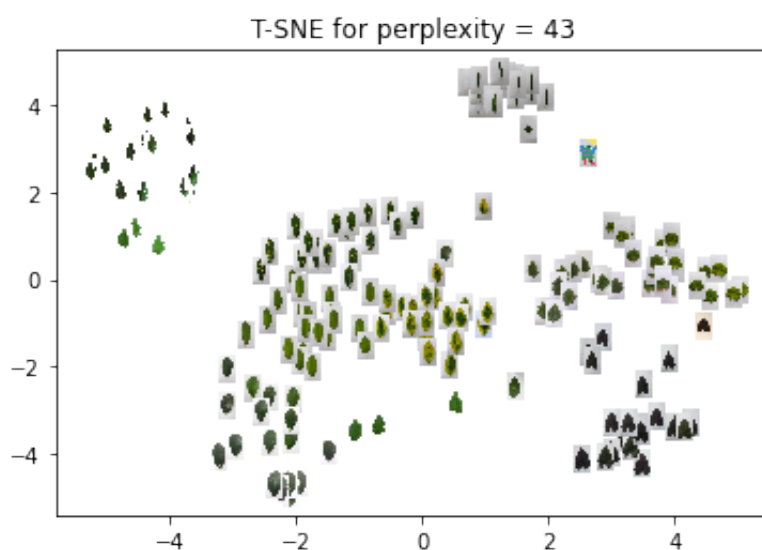
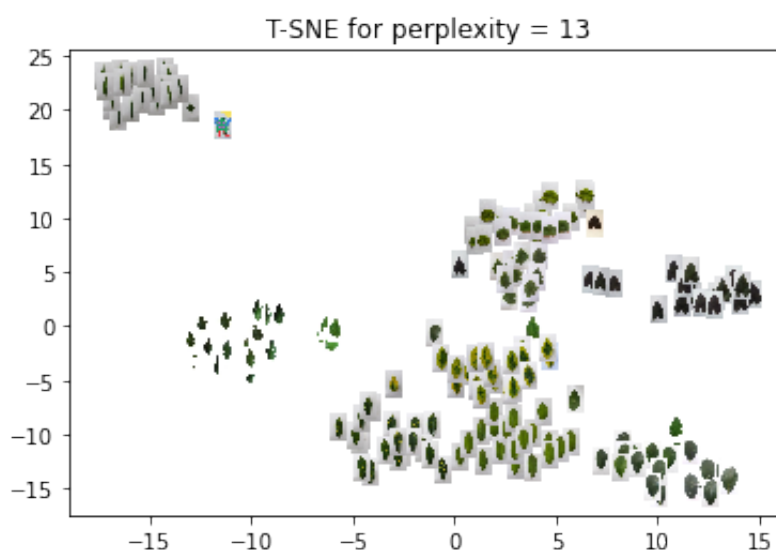


Machine Learning - Project 2

Tâche 1 : T-SNE

J'ai choisi 13 pour la "perplexity" du T-SNE, car on peut voir sur le schéma ci-dessous des zone de hautes densités (notamment en 20-25 sur l'axe des y). Les feuilles sont bien regroupées selon leur forme et on peut déjà apercevoir de futurs clusters (vis quality mesure). Il y a également peu de feuilles orphelines ou mal situées. On peut remarquer que plus on augmente la "perplexity", plus l'algorithme devient "généraliste" et moins les séparations claires (comme par exemple sur la figure 2 avec une "perplexity" de 43). On peut voir néanmoins que certains feuilles restent toujours groupées et forme un ensemble homogène.



Tâche 2 : Clustering avec k-means

J'ai choisi k-means car il est facile à comprendre, efficace et correspond à la division attendue entre les différentes familles de feuilles. J'ai choisi un nombre de cluster égal à 8 car il correspond au nombre de différentes feuilles possibles dans le dataset. Il correspond aussi à l'angle où la SSE a le plus diminué de manière significative (voir image ci-dessous). On peut voir qu'il correspond au résultat obtenu à partir du T-SNE.

