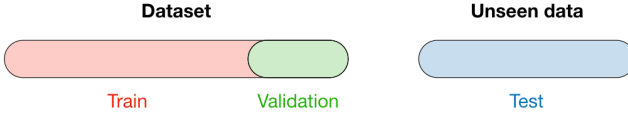


Une fois que le modèle a été choisi, il est entraîné sur le jeu de données entier et testé sur l'ensemble de test (qui n'a jamais été vu). Ces derniers sont représentés dans la figure ci-dessous :



1.6 Apprentissage non supervisé

Les méthodes d'apprentissage non supervisé visent à découvrir la structure (parfois riche) des données.

1.6.1 k -moyennes (en anglais k -means)

Partitionnement – Étant donné un ensemble d'entraînement $\mathcal{D}_{\text{train}}$, le but d'un algorithme de partitionnement (en anglais *clustering*) est d'assigner chaque point $\phi(x_i)$ à une partition $z_i \in \{1, \dots, k\}$.

Fonction objectif – La fonction objectif d'un des principaux algorithmes de partitionnement, k -moyennes, est donné par :

$$\text{Loss}_{k\text{-means}}(x, \mu) = \sum_{i=1}^n \|\phi(x_i) - \mu_{z_i}\|^2$$

Algorithme – Après avoir aléatoirement initialisé les centroïdes de partitions $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$, l'algorithme k -moyennes répète l'étape suivante jusqu'à convergence :

$$z_i = \arg \min_j \|\phi(x_i) - \mu_j\|^2$$

and

$$\mu_j = \frac{\sum_{i=1}^m 1_{\{z_i=j\}} \phi(x_i)}{\sum_{i=1}^m 1_{\{z_i=j\}}}$$

