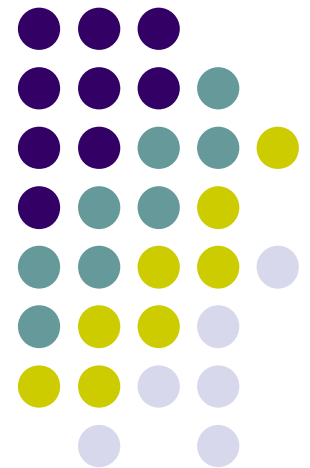
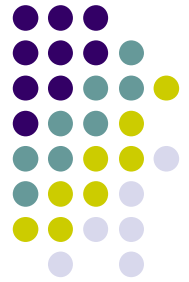


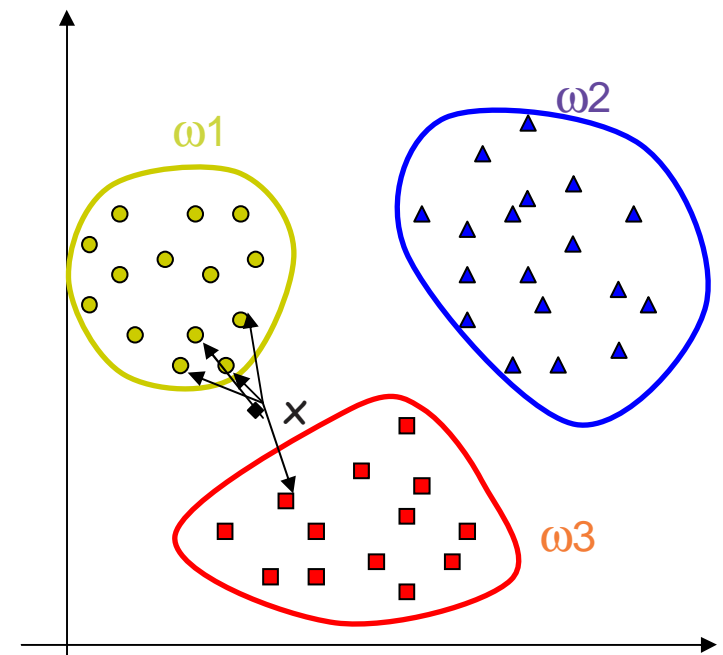
K plus proches voisins



Principe



- **Méthode très intuitive qui classe les exemples non étiquetés sur la base de leur similarité avec les exemples de la base d'apprentissage**
 - Pour un exemple non étiqueté x , trouver les k "plus proches exemples étiquetés de la base d'apprentissage et affecter à x la classe qui apparaît le plus souvent
- **Les k -PPV nécessitent seulement :**
 - Un entier k
 - Une base d'apprentissage
 - Une métrique pour la proximité
- **Exemple**
 - Dans l'exemple suivant, on a 3 classes et le but est de trouver la valeur de la classe de l'exemple inconnu x
 - On prend la distance Euclidienne et $k=5$ voisins
 - Des 5 plus proches voisins, 4 appartiennent à ω_1 et 1 appartient à ω_3 , donc x est affecté à ω_1 , la classe majoritaire





1-Plus proche voisin

$$\text{Dist}(c_1, c_2) = \sqrt{\sum_{i=1}^N (\text{attr}_i(c_1) - \text{attr}_i(c_2))^2}$$

$$\text{PlusProcheVoisin} = \text{MIN}_j (\text{Dist}(c_j, c_{test}))$$

- Marche bien si pas d'attribut ou de classe bruitée

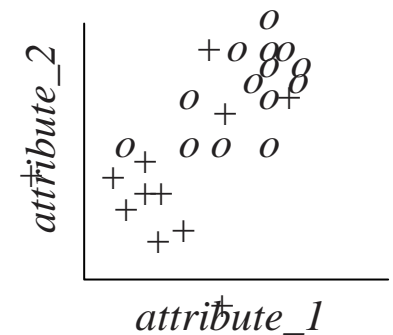


k-Plus proches voisins

$$\text{Dist}(c_1, c_2) = \sqrt{\sum_{i=1}^N (\text{attr}_i(c_1) - \text{attr}_i(c_2))^2}$$

$$k - \text{plus proches voisins} = \{k - \text{MIN}(\text{Dist}(c_i, c_{\text{test}}))\}$$

- Moyenne des k points plus fiable quand:
 - Bruit dans les attributs
 - Bruit dans les valeurs de classe
 - Classes partiellement recouvertes





Comment choisir “k” ?

- K grand:
 - moins sensible au bruit
 - Une grande base d’apprentissage permet une plus grande valeur de k
- K petit :
 - Rend mieux compte de structures fines
 - Nécessaire pour des petites bases d’apprentissage