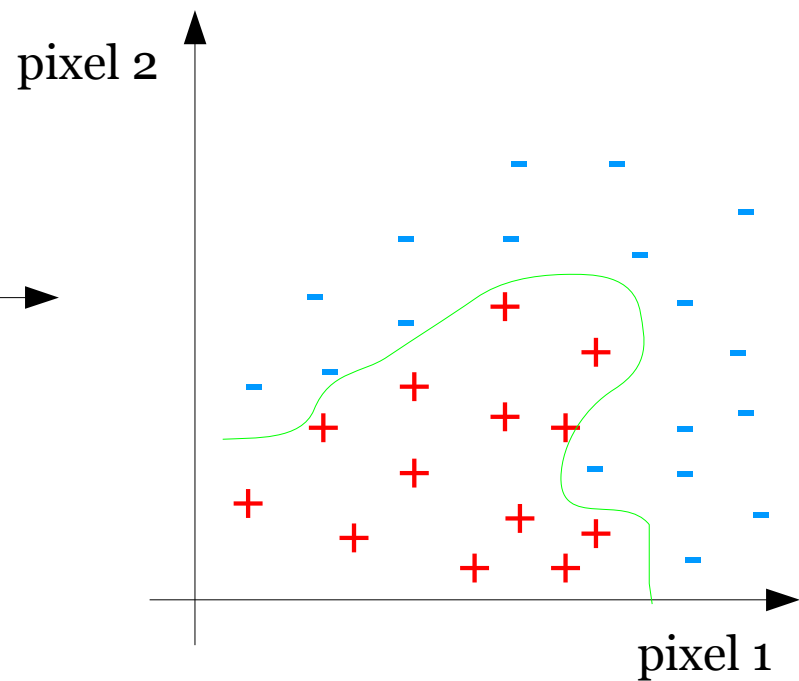
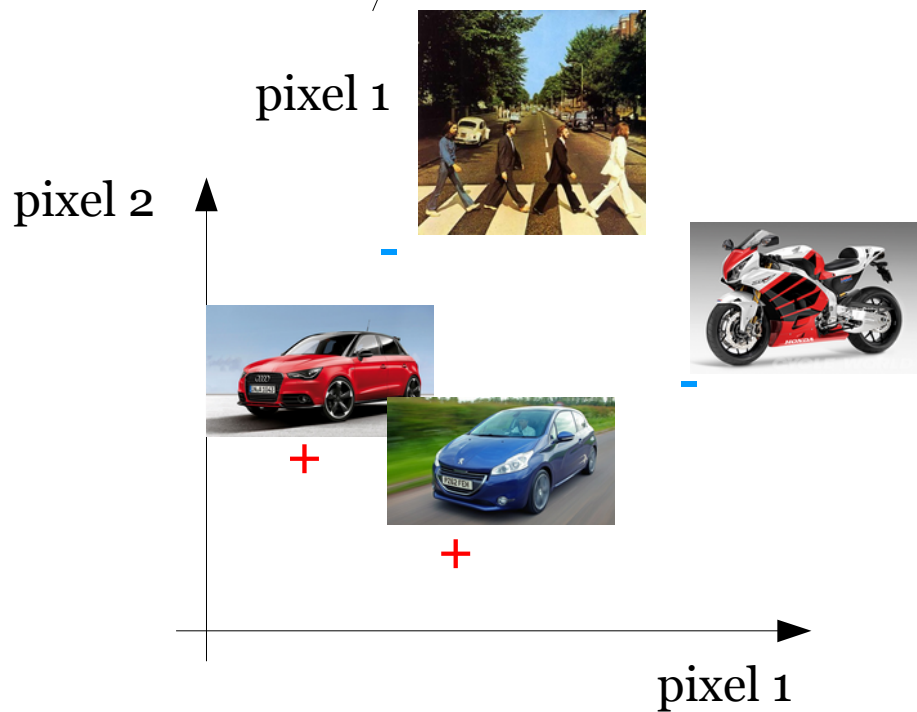


Classification

K-NN



Algorithme d'apprentissage





Voitures



θ

Méthodes paramétriques

- Avantages
 - Classification rapide.
 - Compression de l'information.
- Inconvénients
 - Estimation éloignée du vrai modèle.
 - Apprentissage long et complexe.
 - Apprentissage nécessaire pour l'ajout de nouveaux exemples d'entraînement



Voitures



Quel est cet objet ?



Méthodes non-paramétriques

- Avantages
 - Simple.
 - Pas d'apprentissage nécessaire.
 - Pas de supposition à faire sur l'hypothèse.
- Inconvénients
 - Classification lente.
 - Gourmand en mémoire.
 - Non interprétable.
 - Jeux de données volumineux.

K-NN : k Nearest Neighbors.

K-PPV : k Plus Proches Voisins.

Points importants

- Notion de voisinage
 - Mesure de similarité
- Influence des voisins dans le choix de la classe
 - Nombre de voisins
 - Pondération des voisins
 - Égale pour tous les voisins.
 - Proportionnelle à la distance.
 - Application.

Voisinage

- Mesure de similarité

Observation
Base d'entraînement

		...		
--	--	-----	--	--

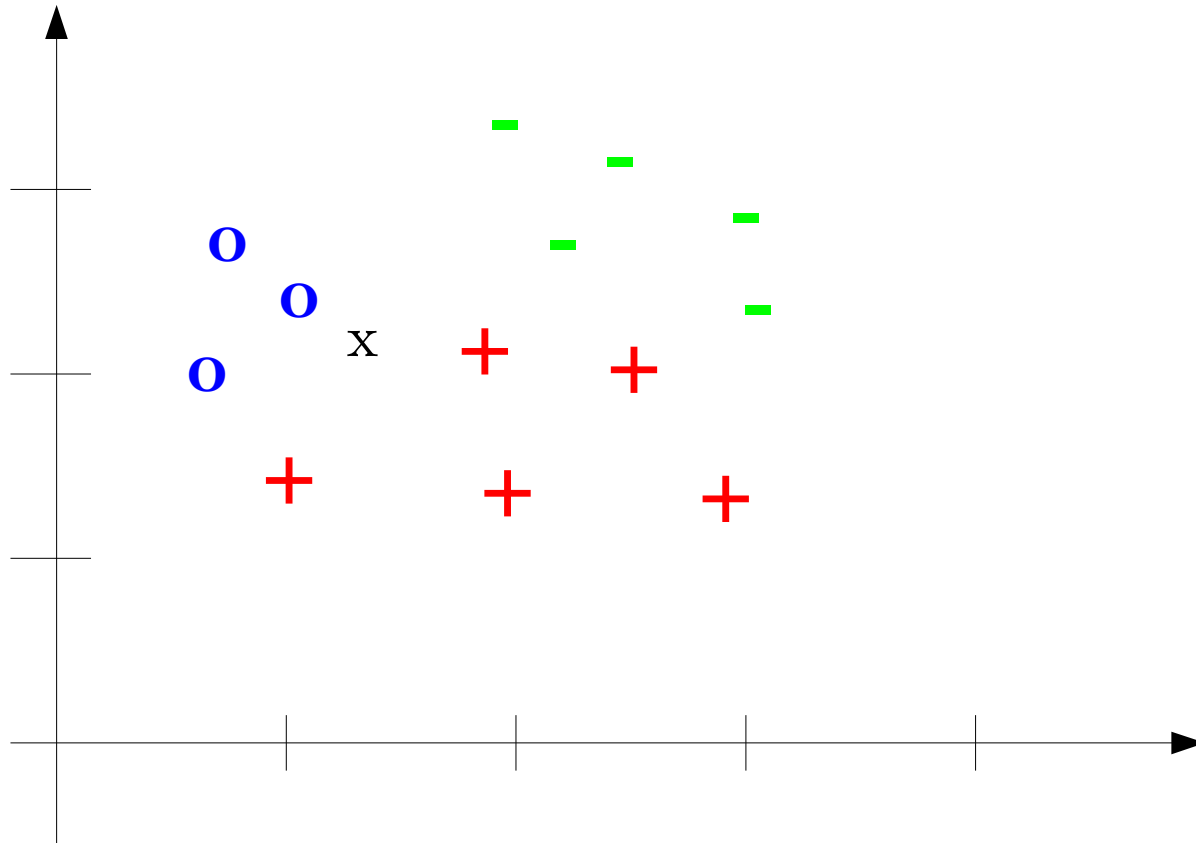
Observation
Candidate

		...		
--	--	-----	--	--

- Distance euclidienne
- Distance de Manhattan
- Distance de Tchebychev
- ...

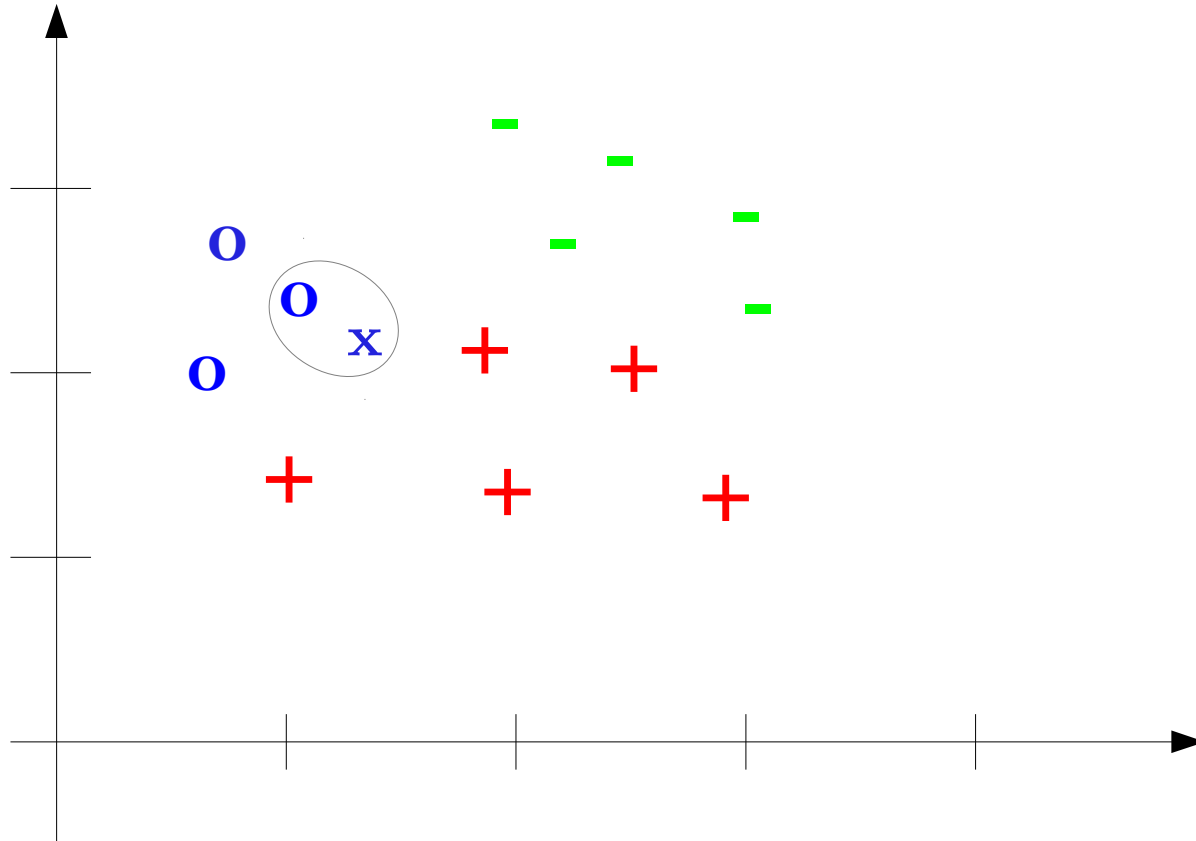
Influence des voisins

- Nombre de voisins



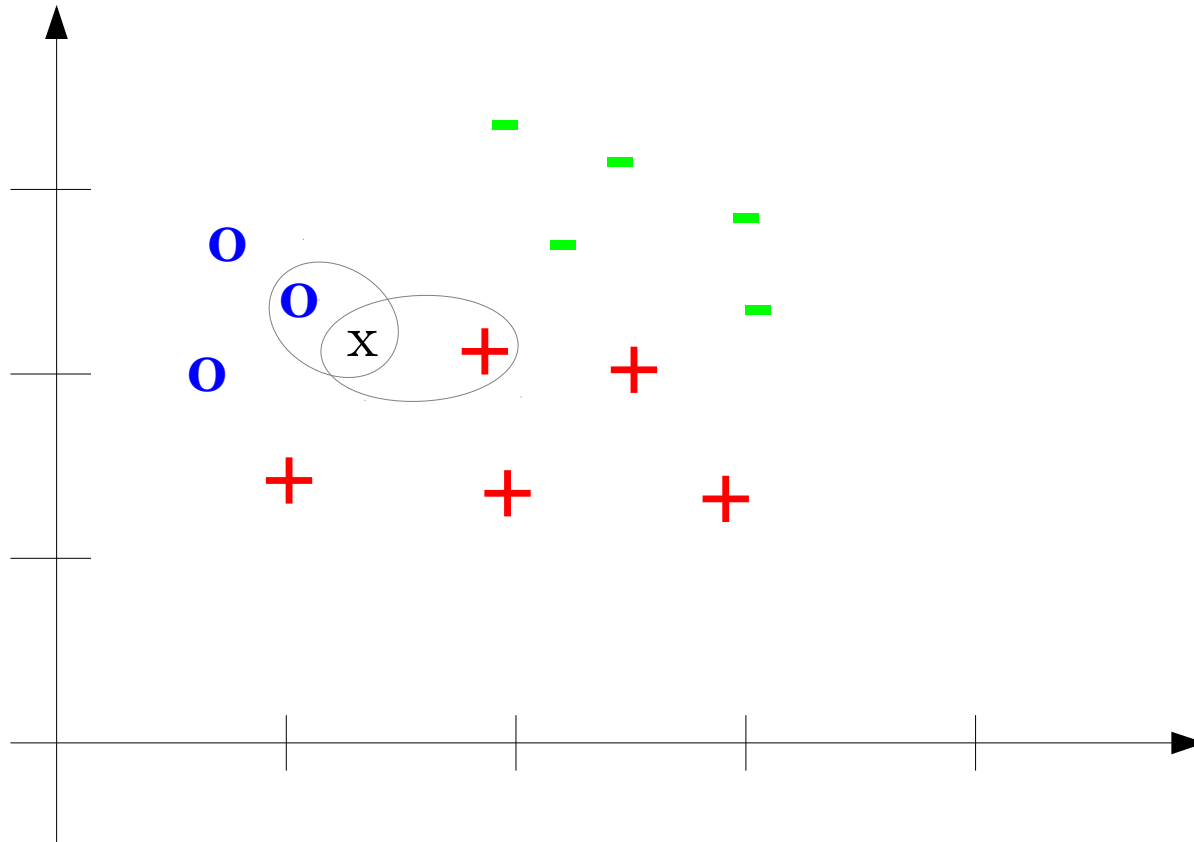
Influence des voisins

- Nombre de voisins
 - 1-PPV



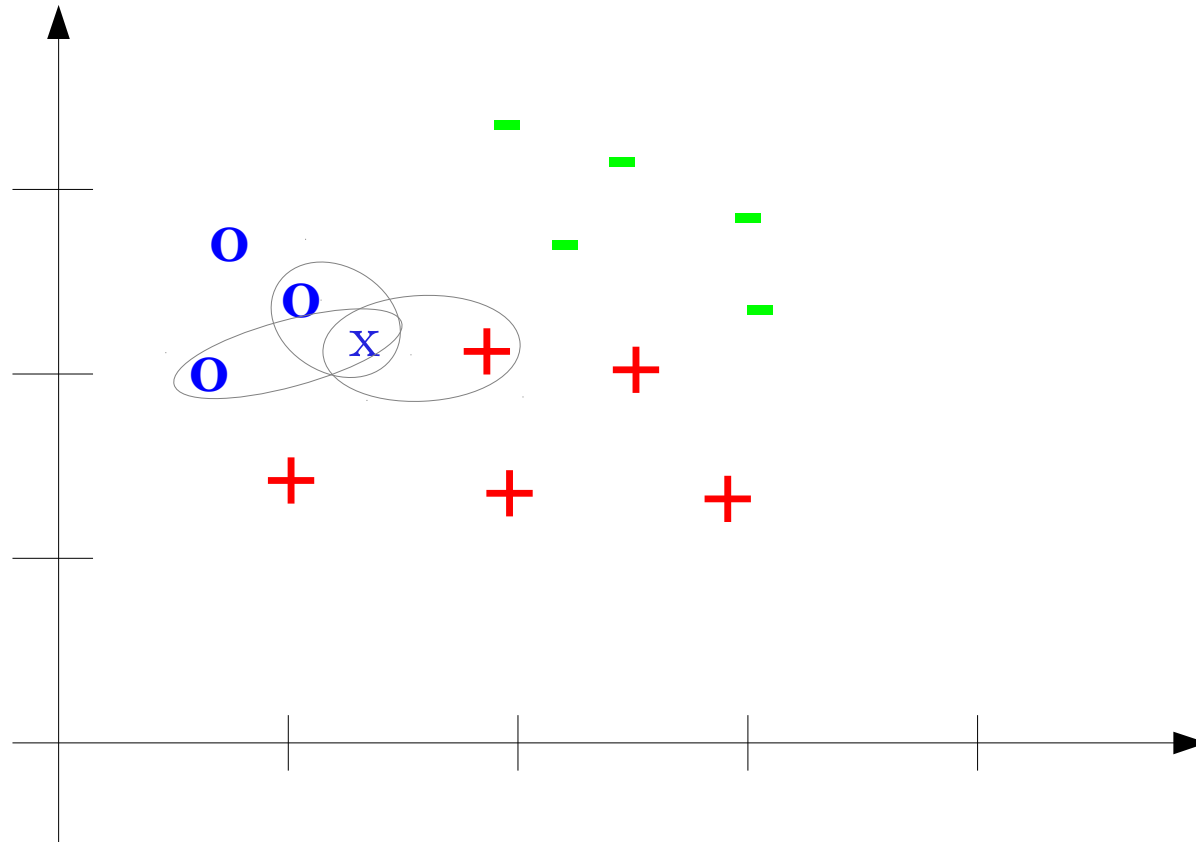
Influence des voisins

- Nombre de voisins
 - 2-PPV

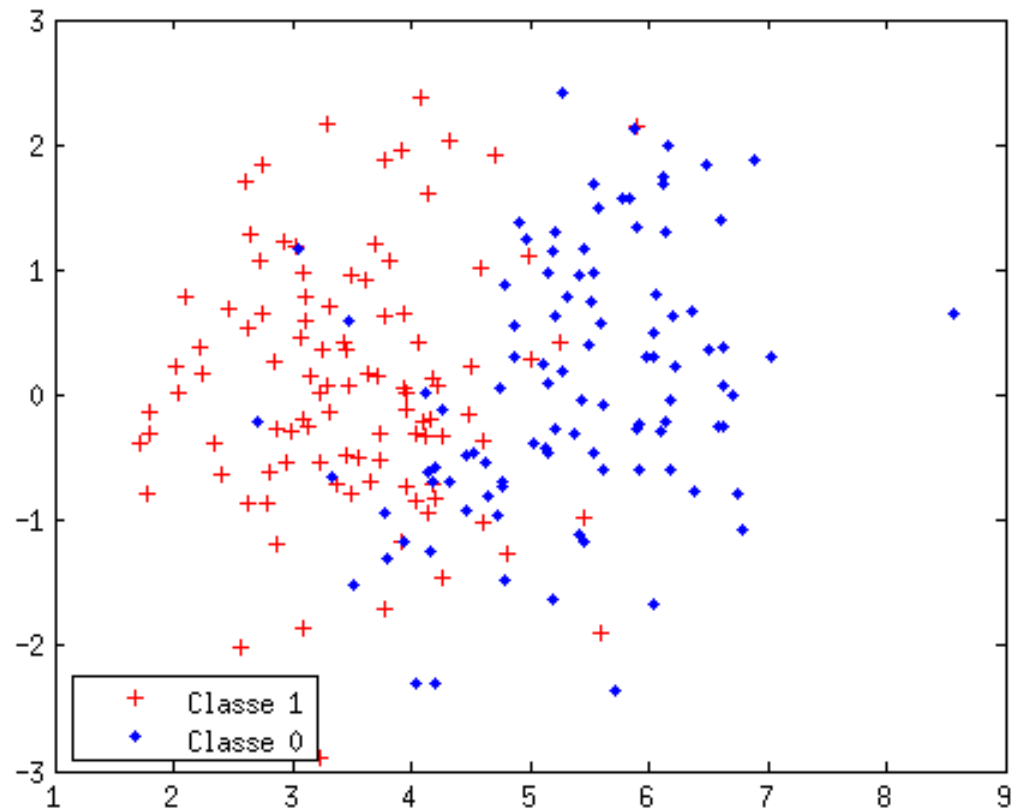


Influence des voisins

- Nombre de voisins
 - 3-PPV

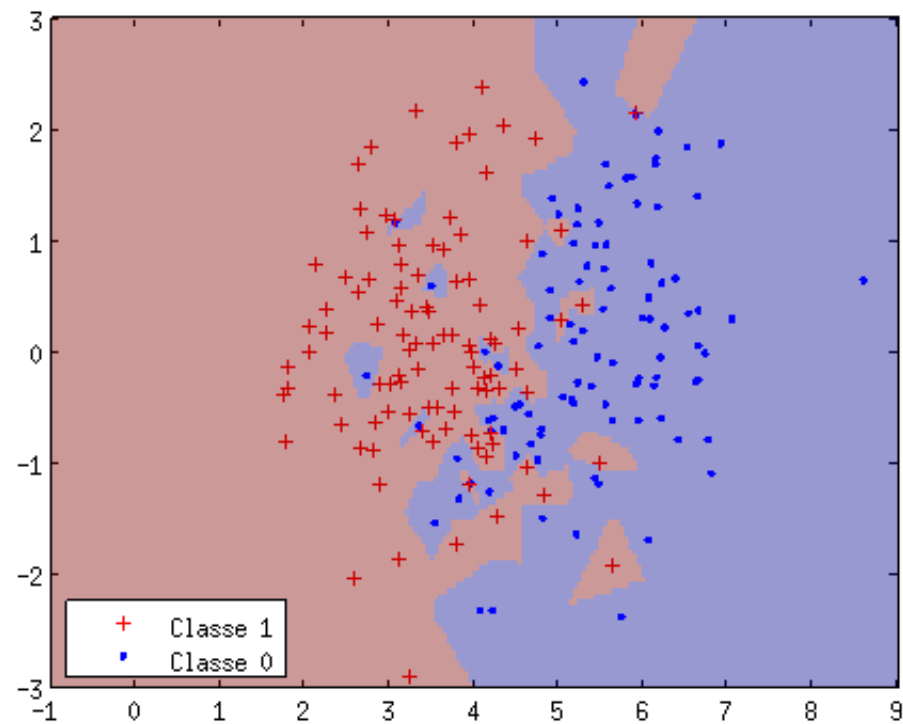


Influence des voisins



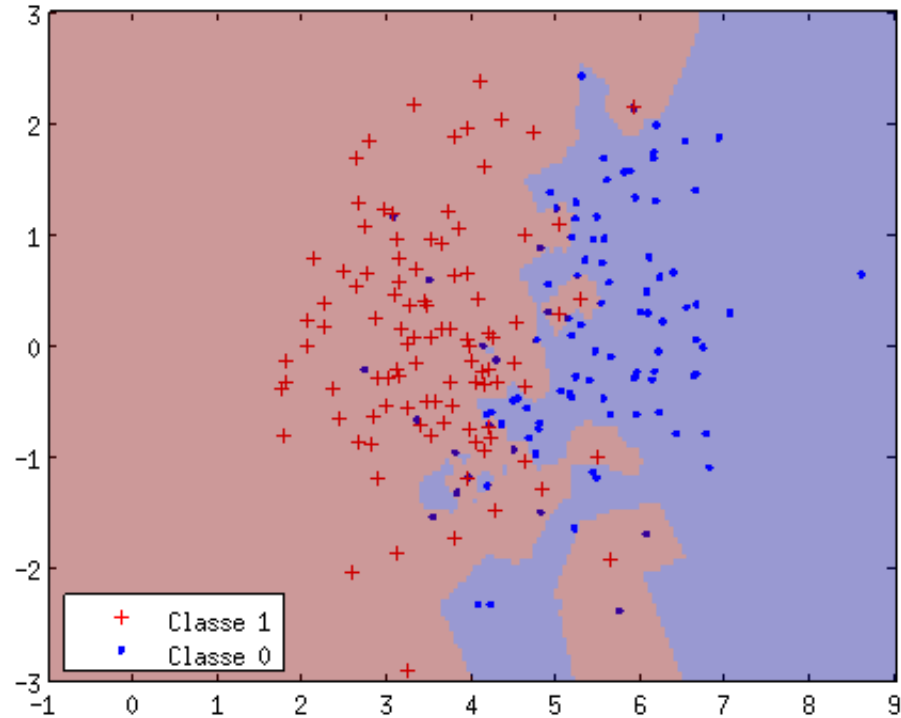
Influence des voisins

- $k=1$



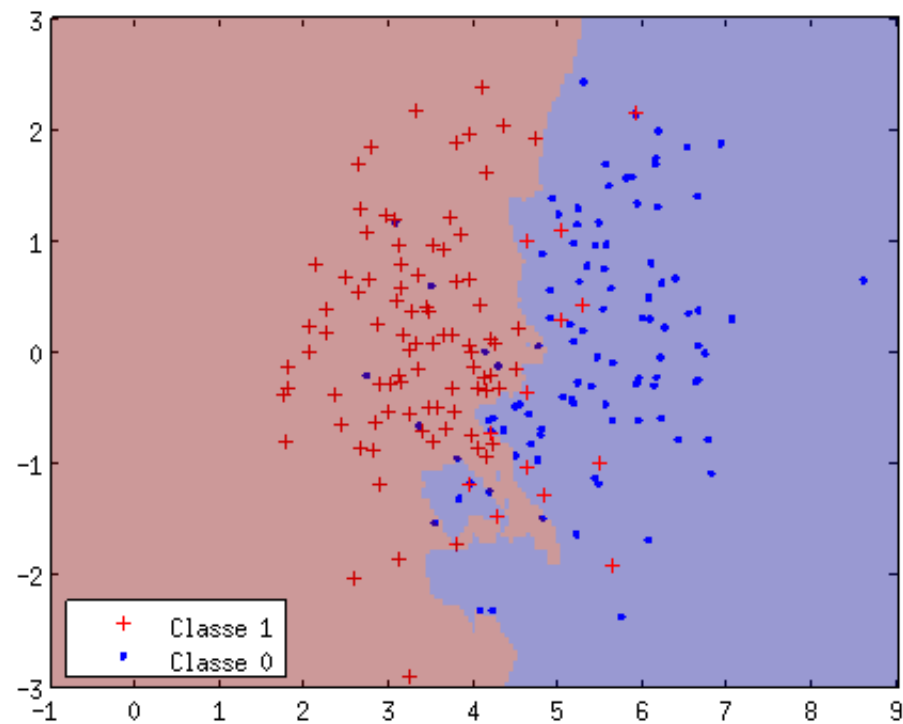
Influence des voisins

- $k=2$



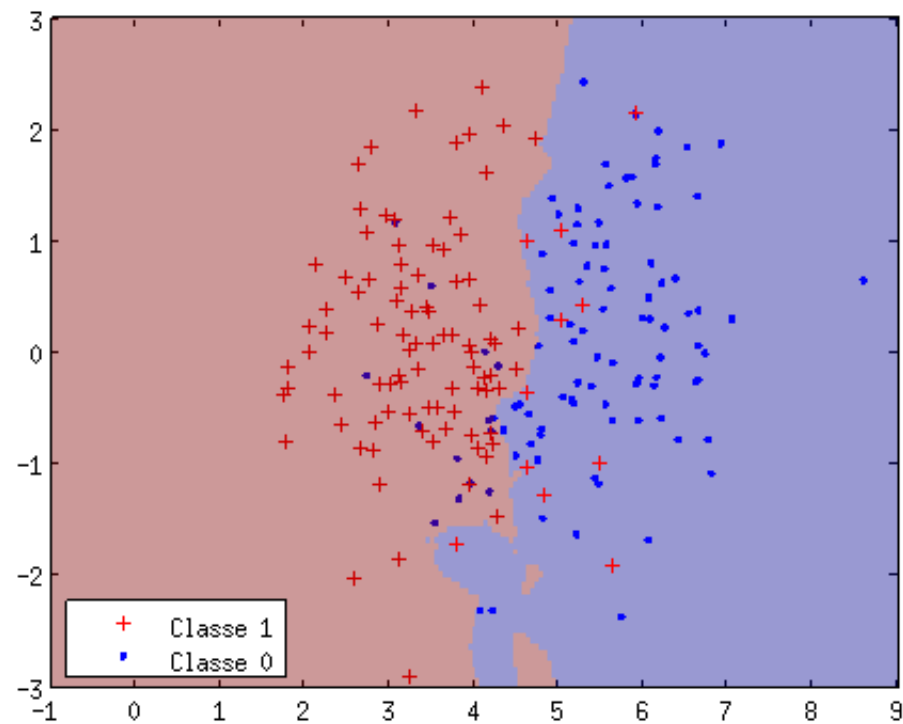
Influence des voisins

- $k=5$



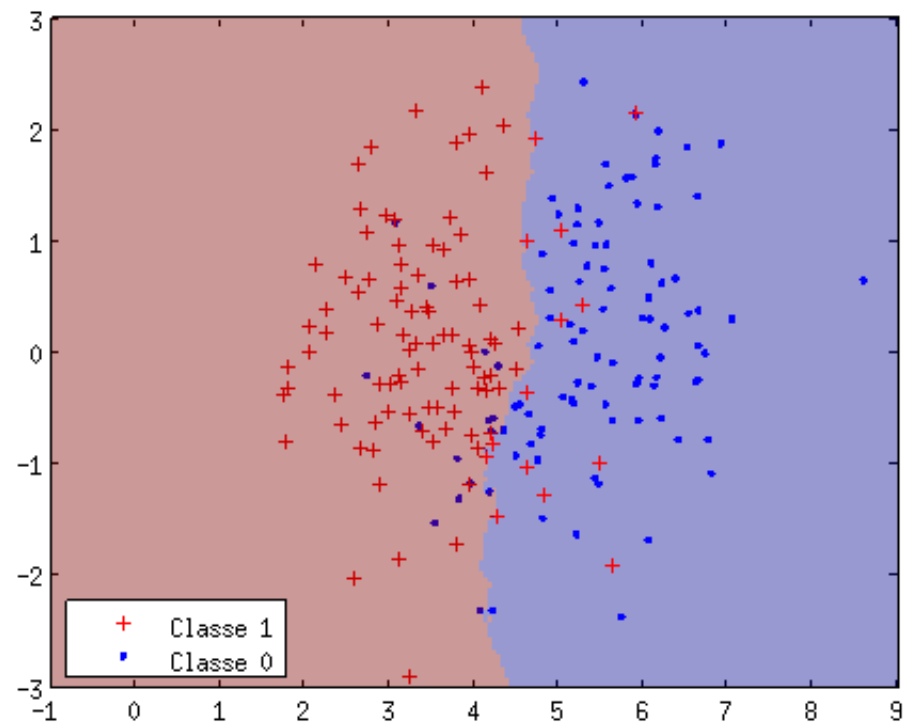
Influence des voisins

- $k=10$



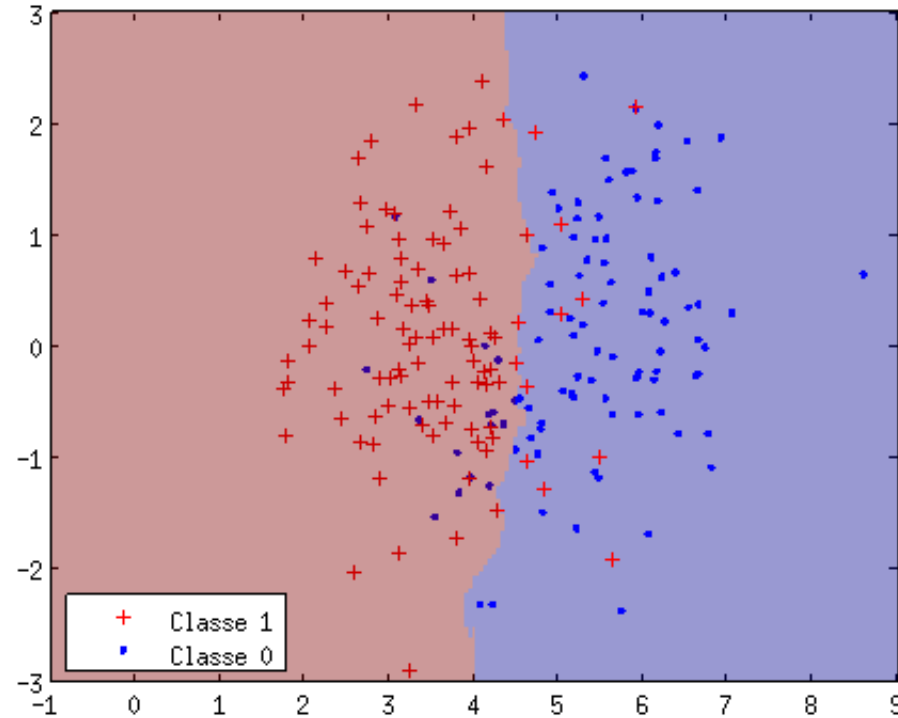
Influence des voisins

- $k=20$



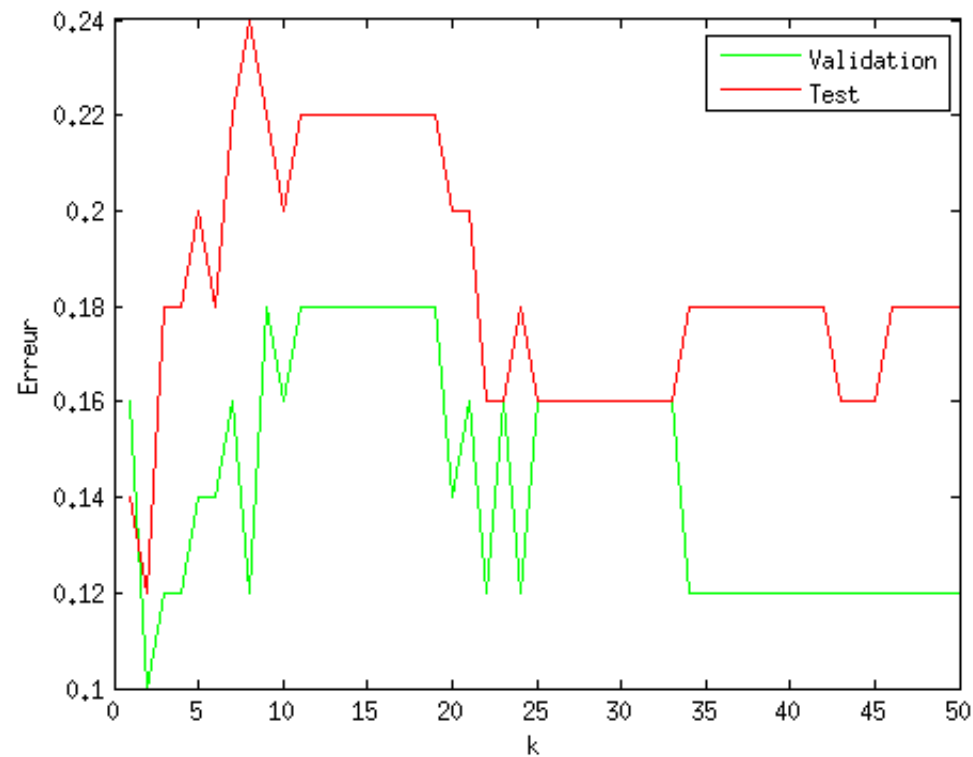
Influence des voisins

- $k=50$



Influence des voisins

- Performance de classification



Merci de votre Attention

Questions

Lab

- Sur les bases « Breast Cancer » et « Heart Disease » :
 - Implémentez l'algorithme des plus proches voisins
 - Évaluez la performance pour $k=1, 2, 5$ et 10 .
 - Normalisez les attributs et comparez avec la performance obtenue précédemment.
 - Comparez vos résultats avec ceux obtenus avec les réseaux de neurones.