

Classifieur k plus proches voisins

- Possiblement l'algorithme d'apprentissage pour la classification le plus simple
- **Idée:** étant donnée une entrée \mathbf{X}
 1. trouver les k entrées \mathbf{X}_t parmi les exemples d'apprentissage qui sont les plus « proches » de \mathbf{X}
 2. faire voter chacune de ces entrées pour leur classe associée y_t
 3. retourner la classe majoritaire
- Le succès de cet algorithme va dépendre de deux facteurs
 - ◆ la quantité de données d'entraînement (plus il y en a, meilleure sera la performance)
 - ◆ la qualité de la mesure de distance (2 entrées similaires sont-elles de la même classe?)
 - » en pratique, on utilise souvent la distance Euclidienne:

$$d(\mathbf{x}_1, \mathbf{x}_2) = \sqrt{\sum_k (x_{1,k} - x_{2,k})^2}$$

\mathbf{x} = vecteur
 x = scalaire

Illustration: 3 plus proches voisins

- Reconnaissance de caractère: est-ce un 'e' ou un 'o'?

Ensemble d'entraînement

(100 exemples d'apprentissage par classe)



Illustration: 3 plus proches voisins

- Reconnaissance de caractère: est-ce un 'e' ou un 'o'?

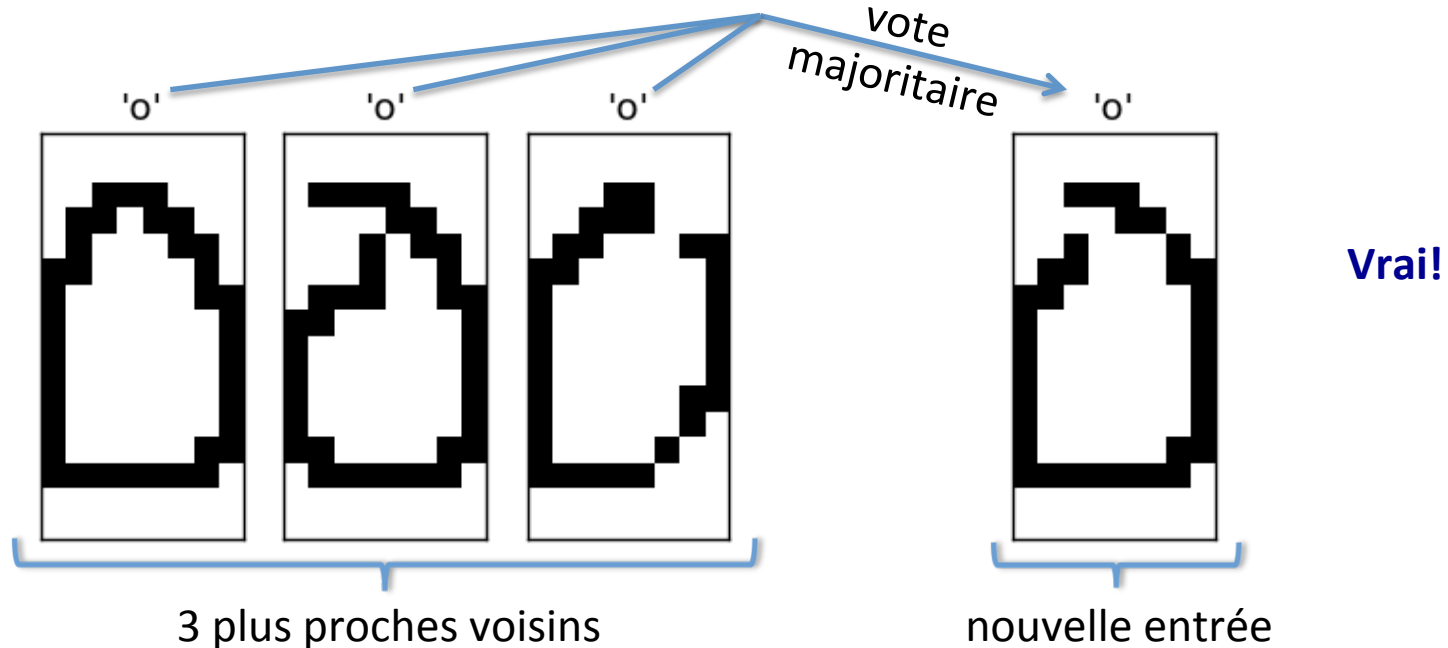


Illustration: 3 plus proches voisins

- Reconnaissance de caractère: est-ce un 'e' ou un 'o'?

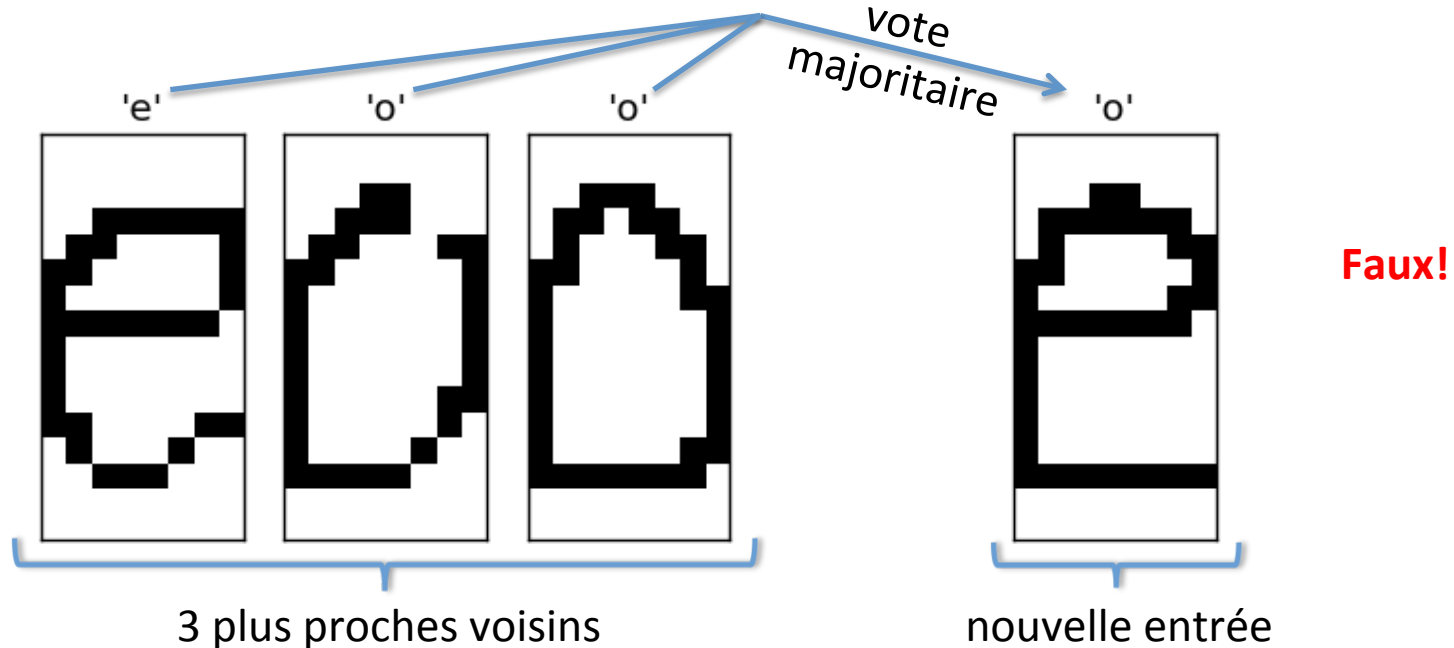
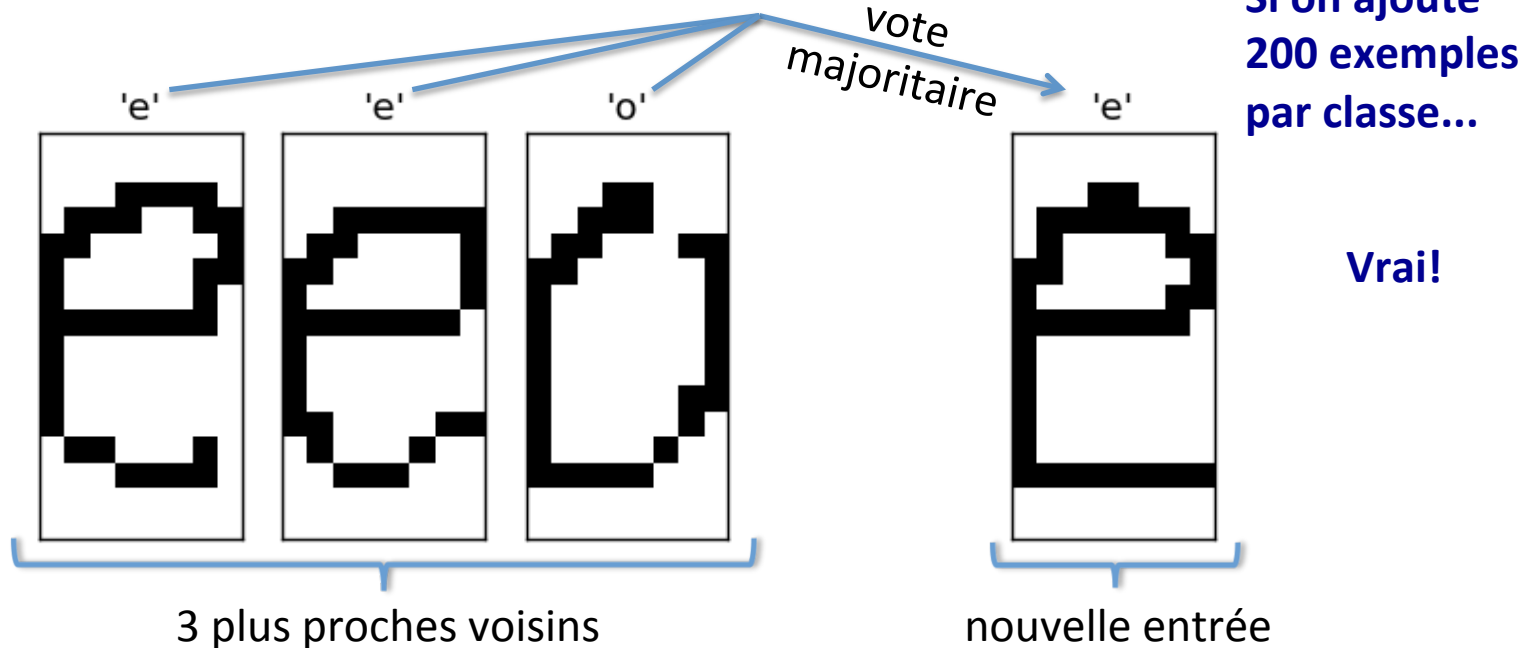


Illustration: 3 plus proches voisins

- Reconnaissance de caractère: est-ce un 'e' ou un 'o'?



Apprentissage supervisé

- Un problème d'apprentissage supervisé est formulé comme suit:
« Étant donné un **ensemble d'entraînement** de N exemples:

$$(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N) \} D$$

où chaque y_j a été généré par une **fonction inconnue** $y = f(\mathbf{x})$,
découvrir une nouvelle fonction h (**modèle** ou **hypothèse**)
qui sera une bonne approximation de f (c'est à dire $f(\mathbf{x}) \approx h(\mathbf{x})$) »

- Un algorithme d'apprentissage peut donc être vu comme étant une fonction A
à laquelle on donne un ensemble d'entraînement et qui donne en retour cette
fonction h

$$A(D) = h$$

Retour sur classifieur k plus proches voisins

- Dans le cas de l'algorithme k plus proches voisins:
 - ◆ A est un programme qui produit lui-même un programme, soit celui qui fait une prédiction à l'aide de la procédure k plus proches voisins
 - ◆ $h = A(D)$ est le programme qui fait voter les k plus proches voisins dans D d'une entrée donnée
 - ◆ $h(\mathbf{x})$ est la sortie du programme pour l'entrée \mathbf{x} , c'est à dire une prédiction de la classe de \mathbf{x}
 - ◆ f est la « fonction » qui a généré nos données d'entraînement
 - » ex.: l'être humain qui a étiqueté les images de caractères
- On peut démontrer que plus D est grand, plus h sera une bonne approximation de f
 - ◆ intuition: en augmentant la taille de l'ensemble d'entraînement, les k plus proches voisins ne peuvent qu'être plus proches (plus similaires) à l'entrée