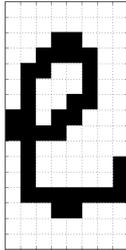


Représentation des données

- L'**entrée** \mathbf{X} est représentée par un vecteur de valeurs d'attributs réels (représentation factorisée)
 - ◆ ex.: une image est représentée par un vecteur contenant la valeur de chacun des pixels



```
array([ 0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  
        0.,  0.,  0.,  0.,  0.,  0.,  1.,  1.,  0.,  0.,  0.,  0.,  0.,  
        1.,  1.,  1.,  1.,  0.,  0.,  0.,  1.,  1.,  0.,  0.,  1.,  0.,  
        0.,  0.,  1.,  0.,  0.,  0.,  1.,  0.,  0.,  0.,  1.,  0.,  0.,  
        1.,  1.,  0.,  0.,  1.,  1.,  0.,  1.,  1.,  0.,  0.,  0.,  1.,  
        1.,  1.,  1.,  0.,  0.,  0.,  0.,  0.,  1.,  0.,  0.,  0.,  0.,  
        0.,  0.,  0.,  1.,  0.,  0.,  0.,  0.,  0.,  1.,  0.,  1.,  0.,  
        0.,  0.,  0.,  0.,  1.,  0.,  1.,  1.,  1.,  1.,  1.,  1.,  1.,  
        0.,  0.,  0.,  1.,  1.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  
        0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.,  0.]
```

- La **sortie désirée** ou **cible** y aura une représentation différente selon le problème à résoudre:
 - ◆ problème de classification en C classes: valeur discrète (index de 0 à $C-1$)
 - ◆ problème de régression: valeur réelle ou continue

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 générée 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

$$A(D) = h$$