

# Rdf - TP5 : Classement de caractère manuscrits

François LEPAN - Benjamin VAN RYSEGHEM

12 mars 2013

## 1 Visualiser les Donnees

```
y = matrix(train0(1,:),28,28)';  
Matplot(y);
```

Afin de visualiser les autres chiffres manuscrits il suffit de changer le nom de train0 en trainN où N est le chiffre à visualiser.

## 2 L'algorithme de plus proche voisin ou kNN

### 2.1 Est-ce que les résultats changent si $d^2(x, y) = \sum_{i=1}^n (x_i - y_i)^2$ est utilisée comme distance ?

Non car tout ce que l'on fait c'est comparer ces distances ( $\geq 0$ ) donc leurs valeurs élevées au carré ne changeront rien lors de la comparaison.

Ex : si a = 2 et b = 3

$$a < b \Rightarrow a^2 < b^2$$

$$2 < 3 \Rightarrow 4 < 9$$

### 2.2 Quelles sont les erreurs moyennes de classification pour k = 1, 3, 5, 10 ?

Pour k = 1 on a une erreur moyenne de 0.6885923.

Étant donné que la plupart des gens ont 0.1 je me dis qu'il doit y avoir une erreur dans notre code mais je n'arrive pas à savoir où.

### 2.3 Cette approche est-elle supervisée ?

Oui cette approche est supervisée, car il faut l'alimenter de données d'entraînement afin de savoir comment classer le nouvel élément entrant.

## 3 L'algorithme Naïve Bayes

### 3.1 Quelle est l'hypothèse principale dans cette approche ?

$$P(X = (x_1, \dots, x_m) \mid C_i) = \prod_{j=1}^m P(x_j \mid C_i)$$

Propriété : L'indépendance conditionnelle des attributs donc on peut faire le produit de ses éléments.

### 3.2 Calcul des erreurs de classification

????????????????

### 3.3 Cette approche est-elle supervisée ?

Oui cette approche est supervisée, car elle utilise des données d'entraînement tout comme kNN.

### 3.4 Cette approche est-elle probabiliste ?

Oui car elle est basée sur la fréquence de chaque valeur (ici  $0 \rightarrow 255$ ) quelque soit le point d'entraînement et quel que soit l'emplacement de cette valeur dans le point.