

# TP KPPV

## 2) Les données

Avec des commandes size successifs, on trouve :

- 1000 échantillons dans la base d'apprentissage
- 300 échantillons dans la base de test

On obtient successivement :

```
for i=1:10
```

```
sum(S==i)
```

```
end
```

116

99

93

105

92

94

117

87

100

97

Ils ne sont pas équirépartis mais la variance est faible.

## 3) Classement par kppv

Pour k=1 :

```
errorRate =
```

0.2167

```
confusion =
```

38	0	0	0	0	0	0	0	0	0
2	20	0	3	1	1	1	0	0	0
1	3	14	0	5	0	0	2	3	0
1	1	0	27	0	1	1	1	6	0
0	1	0	0	17	1	0	2	1	2
1	0	0	0	0	23	0	0	0	1
0	0	0	3	0	0	26	0	4	1
1	0	0	0	3	1	1	23	1	1
0	0	0	1	0	0	3	0	28	0
1	0	0	0	1	0	0	1	0	19

Pour k=3, on a :

```
errorRate =
```

```
0.2067
```

```
confusion =
```

38	0	0	0	0	0	0	0	0	0
3	22	0	1	1	0	1	0	0	0
0	3	13	0	7	0	0	1	4	0
3	4	0	27	1	0	0	0	3	0
1	2	2	1	17	0	0	1	0	0
1	0	0	0	1	23	0	0	0	0
0	0	0	2	1	0	27	0	4	0
2	1	1	0	1	1	1	23	1	0
0	0	0	0	0	0	2	1	29	0
1	0	0	0	2	0	0	0	0	19

Pour k=4, on a :

```
errorRate =
```

```
0.2300
```

```
confusion =
```

38	0	0	0	0	0	0	0	0	0
3	22	0	0	1	0	1	0	0	1
1	1	13	0	7	0	0	2	4	0
3	3	0	27	0	0	0	0	5	0
1	2	3	1	16	0	0	1	0	0
2	0	0	0	0	22	0	0	0	1
0	0	0	0	3	0	27	0	4	0
2	1	1	2	1	1	0	20	3	0
0	0	0	0	0	0	3	1	28	0
1	0	0	0	3	0	0	0	0	18

Pour k=5, on a :

```
errorRate =
```

```
0.2200
```

```
confusion =
```

38	0	0	0	0	0	0	0	0	0
3	22	0	1	1	0	1	0	0	0
1	0	14	0	7	0	0	2	4	0
3	3	0	28	0	0	0	0	4	0
1	1	4	1	16	0	0	1	0	0
2	0	0	0	0	22	0	0	0	1
0	0	0	1	2	0	25	0	6	0
1	1	1	1	1	1	0	22	3	0
0	0	0	0	0	0	2	1	29	0
0	0	0	1	2	0	0	1	0	18

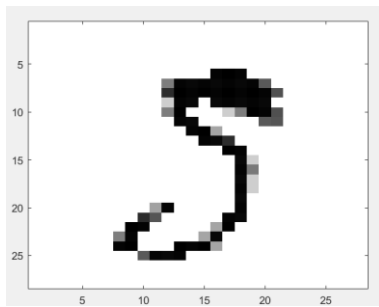
3.3) Dans le cas k=5.

On voit qu'il y a beaucoup de confusion entre le 5 qui est pris pour un 3 : `confusion(5,3)`

Ce sont les samples suivants qui ont été confondus

```
i =  
33  
  
i =  
207  
|  
i =  
211  
  
i =  
223
```

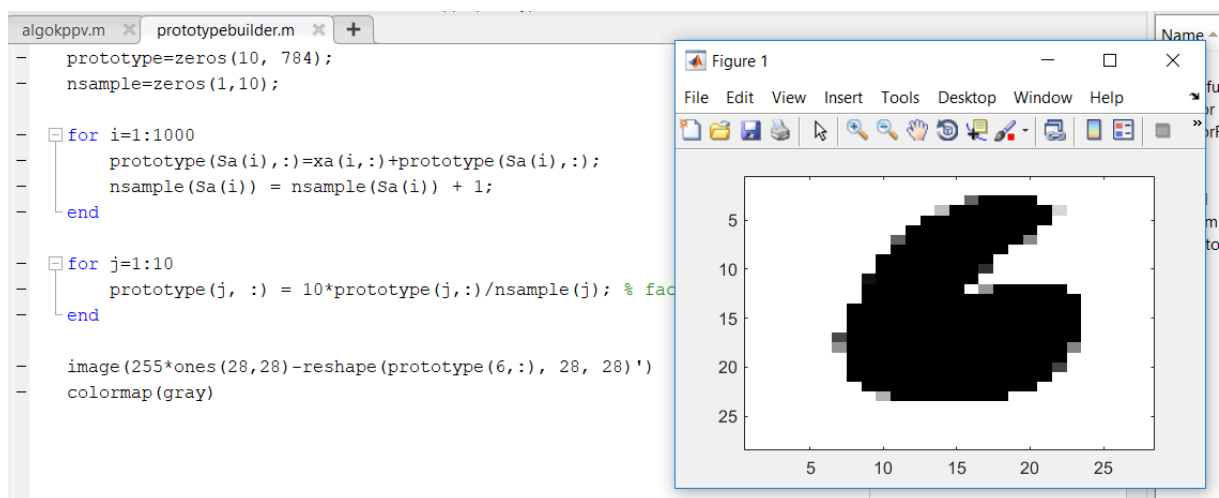
Regardons le sample 33 :



On constate que ce n'est effectivement pas un « beau » 5 car la tête est trop arrondie.

## 4) 1-ppv avec prototypes

4.1) On propose les prototypes de forme moyenne. Ici on a la moyenne du 6 :



4.2) On classe :

```
confusion=zeros(10,10);
for i=1:300
    mmin=norm(prototype(1,:)-x(i,:));
    imin=1;
    for j=2:10
        if norm(prototype(j,:)-x(i,:)) < mmin
            mmin=norm(prototype(j,:)-x(i,:));
            imin = j;
        end
    end
    confusion(S(i),imin)=confusion(S(i),imin)+1;
end
```