Series 2a. Mean - Variance Feature - extraction Pattern Recognition

Roberto Sanchez
Univesité de Fribourg

March 28, 2016

1 Description

1.1 Feature Extraction

The script (1.2.Transf.py) reduce the vector of 28 x 28 to a smaller vector. The script **1.2.Transf.py** reduces the size vector to 28 x 2 values. The original vector is split in 28 groups where the colors values (0-255 gray scale) are not important anymore, but instead the position where each pixel has a value greater than the parameter opacity:

$$\mu_i = \frac{1}{N} \sum_{i=0}^{28} X_{i|v>opacity} \tag{1}$$

N is number of pixels that have values of the gray scale greater than *opacity* parameter.
In addition, the variance is also calculate:

$$\sigma_i = \frac{1}{N} \left[\sum_{i=0}^{28} (X_{i|v>opacity} - \mu)^2 \right]^{1/2} \tag{2}$$

The final structure of the vector is as follow:

x[0]: Contains the digit handwriting data label.

n : Is number of 28 formed groups, therefore $n = [0...27]$

x[2n+1]: Contains the mean position of the pixels inside of the group of 28.

x[2n+2]: Contains the rest of mean position and the variance: $\mu - \sigma$