# Prediction of Handwritten Characters using SVM

## *1. Introduction*

The task is to train a SVM to correctly classify handwritten characters fed under the form of 16x8 bitmap images. The training set is composed of 41721 examples and the test set on which the predictions are to be made is formed by 10431 inputs.
The images are composed of one feature per pixel (either 0 or 1, with 1→'pixel on') that represent one of the 26 possible lower-case letters of the alphabet

## *2. Procedure*

Since it's a multiclass-classification task, a single SupportVectorMachine couldn't do the whole work by itself, so it has been used the “one-vs-all” approach. The problem is tackled training a binary classifier per class setting the i-th class for the i-th classifier as the class “1” and the others as “0” effectively instructing a single classifier to focus on a single possible outcome. The prediction is chosen as the highest confidence among the result of all SVMs given a certain input (confidence intended as the value of the function $f_i(\mathbf{x}) = \mathbf{w}_i^T\mathbf{x} + w_{0,i}$).

***Parameters:***

- *Kernels*: Gaussian, linear
- *C*: 1, 10, 100, 1000
- *gamma* = 0.1, 0.01, 0.001, 0.0001

$$(1)\quad min_w \frac{1}{2}\|w\|^2 + C\sum l(y_i, f(x_i))$$

$$(2)\quad k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}'\|^2}{2\sigma^2}\right)$$

***C*** is the normalization constant that trades off the relevance of the classification and margin errors with the width of the separation margin (1), ***gamma*** is the value of $\sigma^2$ for the Gaussian kernel that tweaks the “width” of the decision boundary (2).

# 3. Analysis

| C | Gamma | Precision | Recall | F1 |
|---|---|---|---|---|
| 1 | 0.1 | 0.887027826312 | 0.883583789856 | 0.881752437836 |
| 1 | 0.01 | 0.840005937659 | 0.835071052682 | 0.833277134399 |
| 1 | 0.001 | 0.750717037644 | 0.744972525475 | 0.732405734865 |
| 10 | 0.1 | 0.890638100303 | 0.888569254305 | 0.887370859318 |
| 10 | 0.01 | 0.881899772397 | 0.879748705666 | 0.879010306665 |
| 10 | 0.001 | 0.823298848525 | 0.818101174471 | 0.816438196595 |
| 100 | 0.1 | 0.889700388532 | 0.887850183919 | 0.886617762475 |
| 100 | 0.01 | 0.887076822568 | 0.886531866763 | 0.886041841281 |
| 100 | 0.001 | 0.841224090782 | 0.838546406358 | 0.83741376311 |

The table above describes some of the cross-validation scores obtained with the Gaussian kernel.
The best configuration found was the one with C=10 and gamma=0.1.

The learning curve show below shows how the behaviour of the SVM during the learning