

Support Vector Machines for Non-Linearly Separable Data

Prev Tutorial: [Introduction to Support Vector Machines](#)

Next Tutorial: [Introduction to Principal Component Analysis \(PCA\)](#)

Original author	Fernando Iglesias García
Compatibility	OpenCV >= 3.0

Goal

In this tutorial you will learn how to:

- Define the optimization problem for SVMs when it is not possible to separate linearly the training data.
- How to configure the parameters to adapt your SVM for this class of problems.

Motivation

Why is it interesting to extend the SVM optimization problem in order to handle non-linearly separable training data? Most of the applications in which SVMs are used in computer vision require a more powerful tool than a simple linear classifier. This stems from the fact that in these tasks **the training data can be rarely separated using an hyperplane**.

Consider one of these tasks, for example, face detection. The training data in this case is composed by a set of images that are faces and another set of images that are non-faces (*every other thing in the world except from faces*). This training data is too complex so as to find a representation of each sample (*feature vector*) that could make the whole set of faces linearly separable from the whole set of non-faces.

Extension of the Optimization Problem

Remember that using SVMs we obtain a separating hyperplane. Therefore, since the training data is now non-linearly separable, we must admit that the hyperplane found will misclassify some of the samples. This *misclassification* is a new variable in the optimization that must be taken into account. The new model has to include both the old requirement of finding the hyperplane that gives the biggest margin and the new one of generalizing the training data correctly by not allowing too many classification errors.

We start here from the formulation of the optimization problem of finding the hyperplane which maximizes the **margin** (this is explained in the previous tutorial ([Introduction to Support Vector Machines](#))):

$$\min_{\beta, \beta_0} L(\beta) = \frac{1}{2} \|\beta\|^2 \text{ subject to } y_i(\beta^T x_i + \beta_0) \geq 1 \quad \forall i$$

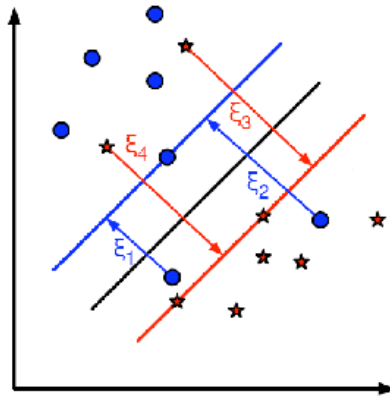
There are multiple ways in which this model can be modified so it takes into account the misclassification errors. For example, one could think of minimizing the same quantity plus a constant times the number of misclassification errors in the training data, i.e.:

$$\min \|\beta\|^2 + C(\text{misclassification errors})$$

However, this one is not a very good solution since, among some other reasons, we do not distinguish between samples that are misclassified with a small distance to their appropriate decision region or samples that are not. Therefore, a better solution will take into account the *distance of the misclassified samples to their correct decision regions*, i.e.:

$$\min \|\beta\|^2 + C(\text{distance of misclassified samples to their correct regions})$$

For each sample of the training data a new parameter ξ_i is defined. Each one of these parameters contains the distance from its corresponding training sample to their correct decision region. The following picture shows non-linearly separable training data from two classes, a separating hyperplane and the distances to their correct regions of the samples that are misclassified.



Note

Only the distances of the samples that are misclassified are shown in the picture. The distances of the rest of the samples are zero since they lay already in their correct decision region.

The red and blue lines that appear on the picture are the margins to each one of the decision regions. It is very **important** to realize that each of the ξ_i goes from a misclassified training sample to the margin of its appropriate region.

Finally, the new formulation for the optimization problem is:

$$\min_{\beta, \beta_0} L(\beta) = \|\beta\|^2 + C \sum_i \xi_i \text{ subject to } y_i(\beta^T x_i + \beta_0) \geq 1 - \xi_i \text{ and } \xi_i \geq 0 \forall i$$

How should the parameter C be chosen? It is obvious that the answer to this question depends on how the training data is distributed. Although there is no general answer, it is useful to take into account these rules:

- Large values of C give solutions with *less misclassification errors* but a *smaller margin*. Consider that in this case it is expensive to make misclassification errors. Since the aim of the optimization is to minimize the argument, few misclassifications errors are allowed.
- Small values of C give solutions with *bigger margin* and *more classification errors*. In this case the minimization does not consider that much the term of the sum so it focuses more on finding a hyperplane with big margin.

Source Code

C++ Java Python

You may also find the source code in `samples/cpp/tutorial_code/ml/non_linear_svms` folder of the OpenCV source library or [download it from here](#).

- **Downloadable code:** Click [here](#)
- **Code at glance:**

```
from __future__ import print_function
import cv2 as cv
import numpy as np
import random as rng

NTRAINING_SAMPLES = 100 # Number of training samples per class
FRAC_LINEAR_SEP = 0.9 # Fraction of samples which compose the linear separable part

# Data for visual representation
WIDTH = 512
HEIGHT = 512
I = np.zeros((HEIGHT, WIDTH, 3), dtype=np.uint8)

# ----- 1. Set up training data randomly -----
trainData = np.empty((2*NTRAINING_SAMPLES, 2), dtype=np.float32)
labels = np.empty((2*NTRAINING_SAMPLES, 1), dtype=np.int32)

rng.seed(100) # Random value generation class

# Set up the linearly separable part of the training data
nLinearSamples = int(FRAC_LINEAR_SEP * NTRAINING_SAMPLES)

trainClass = trainData[0:nLinearSamples,:]
# The x coordinate of the points is in [0, 0.4)
c = trainClass[:,0:1]
c[:] = np.random.uniform(0.0, 0.4 * WIDTH, c.shape)
# The y coordinate of the points is in [0, 1)
```

```

c = trainClass[:,1:2]
c[:] = np.random.uniform(0.0, HEIGHT, c.shape)

# Generate random points for the class 2
trainClass = trainData[2*NTRAINING_SAMPLES-nLinearSamples:2*NTRAINING_SAMPLES,:]
# The x coordinate of the points is in [0.6, 1]
c = trainClass[:,0:1]
c[:] = np.random.uniform(0.6*WIDTH, WIDTH, c.shape)
# The y coordinate of the points is in [0, 1)
c = trainClass[:,1:2]
c[:] = np.random.uniform(0.0, HEIGHT, c.shape)

#----- Set up the non-linearly separable part of the training data -----

trainClass = trainData[nLinearSamples:2*NTRAINING_SAMPLES-nLinearSamples,:]
# The x coordinate of the points is in [0.4, 0.6)
c = trainClass[:,0:1]
c[:] = np.random.uniform(0.4*WIDTH, 0.6*WIDTH, c.shape)
# The y coordinate of the points is in [0, 1)
c = trainClass[:,1:2]
c[:] = np.random.uniform(0.0, HEIGHT, c.shape)

#----- Set up the labels for the classes -----
labels[0:NTRAINING_SAMPLES:] = 1 # Class 1
labels[NTRAINING_SAMPLES:2*NTRAINING_SAMPLES:] = 2 # Class 2

#----- 2. Set up the support vector machines parameters -----
print('Starting training process')

svm = cv.ml.SVM_create()
svm.setType(cv.ml.SVM_C_SVC)
svm.setC(0.1)
svm.setKernel(cv.ml.SVM_LINEAR)
svm.setTermCriteria((cv.TERM_CRITERIA_MAX_ITER, int(1e7), 1e-6))

#----- 3. Train the svm -----

svm.train(trainData, cv.ml.ROW_SAMPLE, labels)

print('Finished training process')

#----- 4. Show the decision regions -----

green = (0,100,0)
blue = (100,0,0)
for i in range(I.shape[0]):
    for j in range(I.shape[1]):
        sampleMat = np.matrix([[j,i]], dtype=np.float32)
        response = svm.predict(sampleMat)[1]

        if response == 1:
            I[i,j] = green
        elif response == 2:
            I[i,j] = blue

#----- 5. Show the training data -----

thick = -1
# Class 1
for i in range(NTRAINING_SAMPLES):
    px = trainData[i,0]
    py = trainData[i,1]
    cv.circle(I, (int(px), int(py)), 3, (0, 255, 0), thick)

# Class 2
for i in range(NTRAINING_SAMPLES, 2*NTRAINING_SAMPLES):
    px = trainData[i,0]
    py = trainData[i,1]
    cv.circle(I, (int(px), int(py)), 3, (255, 0, 0), thick)

#----- 6. Show support vectors -----

thick = 2
sv = svm.getUncompressedSupportVectors()

for i in range(sv.shape[0]):
    cv.circle(I, (int(sv[i,0]), int(sv[i,1])), 6, (128, 128, 128), thick)

```

```
cv.imwrite('result.png', I) # save the Image
cv.imshow('SVM for Non-Linear Training Data', I) # show it to the user
cv.waitKey()
```

Explanation

C++ Java Python

• Set up the training data

The training data of this exercise is formed by a set of labeled 2D-points that belong to one of two different classes. To make the exercise more appealing, the training data is generated randomly using a uniform probability density functions (PDFs).

We have divided the generation of the training data into two main parts.

In the first part we generate data for both classes that is linearly separable.

```
# Generate random points for the class 1
trainClass = trainData[0:nLinearSamples,:]
# The x coordinate of the points is in [0, 0.4]
c = trainClass[:,0:1]
c[:] = np.random.uniform(0.0, 0.4 * WIDTH, c.shape)
# The y coordinate of the points is in [0, 1]
c = trainClass[:,1:2]
c[:] = np.random.uniform(0.0, HEIGHT, c.shape)

# Generate random points for the class 2
trainClass = trainData[2*NTRAINING_SAMPLES-nLinearSamples:2*NTRAINING_SAMPLES,:]
# The x coordinate of the points is in [0.6, 1]
c = trainClass[:,0:1]
c[:] = np.random.uniform(0.6*WIDTH, WIDTH, c.shape)
# The y coordinate of the points is in [0, 1]
c = trainClass[:,1:2]
c[:] = np.random.uniform(0.0, HEIGHT, c.shape)
```

In the second part we create data for both classes that is non-linearly separable, data that overlaps.

```
# Generate random points for the classes 1 and 2
trainClass = trainData[nLinearSamples:2*NTRAINING_SAMPLES,:]
# The x coordinate of the points is in [0.4, 0.6]
c = trainClass[:,0:1]
c[:] = np.random.uniform(0.4*WIDTH, 0.6*WIDTH, c.shape)
# The y coordinate of the points is in [0, 1]
c = trainClass[:,1:2]
c[:] = np.random.uniform(0.0, HEIGHT, c.shape)
```

• Set up SVM's parameters

Note

In the previous tutorial [Introduction to Support Vector Machines](#) there is an explanation of the attributes of the class `cv::ml::SVM` that we configure here before training the SVM.

```
svm = cv.ml.SVM_create()
svm.setType(cv.ml.SVM_C_SVC)
svm.setC(0.1)
svm.setKernel(cv.ml.SVM_LINEAR)
svm.setTermCriteria((cv.TERM_CRITERIA_MAX_ITER, int(1e7), 1e-6))
```

There are just two differences between the configuration we do here and the one that was done in the previous tutorial ([Introduction to Support Vector Machines](#)) that we use as reference.

- C. We chose here a small value of this parameter in order not to punish too much the misclassification errors in the optimization. The idea of doing this stems from the will of obtaining a solution close to the one intuitively expected. However, we recommend to get a better insight of the problem by making adjustments to this parameter.

Note

In this case there are just very few points in the overlapping region between classes. By giving a smaller value to `FRAC_LINEAR_SEP` the density of points can be incremented and the impact of the parameter C explored deeply.

- *Termination Criteria of the algorithm.* The maximum number of iterations has to be increased considerably in order to solve correctly a problem with non-linearly separable training data. In particular, we have increased in five orders of magnitude this value.

• Train the SVM

We call the method `cv::ml::SVM::train` to build the SVM model. Watch out that the training process may take a quite long time. Have patience when you run the program.

```
svm.train(trainData, cv.ml.ROW_SAMPLE, labels)
```

- **Show the Decision Regions**

The method `cv::ml::SVM::predict` is used to classify an input sample using a trained SVM. In this example we have used this method in order to color the space depending on the prediction done by the SVM. In other words, an image is traversed interpreting its pixels as points of the Cartesian plane. Each of the points is colored depending on the class predicted by the SVM; in dark green if it is the class with label 1 and in dark blue if it is the class with label 2.

```
green = (0,100,0)
blue = (100,0,0)
for i in range(I.shape[0]):
    for j in range(I.shape[1]):
        sampleMat = np.matrix([[j,i]], dtype=np.float32)
        response = svm.predict(sampleMat)[1]

        if response == 1:
            I[i,j] = green
        elif response == 2:
            I[i,j] = blue
```

- **Show the training data**

The method `cv::circle` is used to show the samples that compose the training data. The samples of the class labeled with 1 are shown in light green and in light blue the samples of the class labeled with 2.

```
thick = -1
# Class 1
for i in range(NTRAINING_SAMPLES):
    px = trainData[i,0]
    py = trainData[i,1]
    cv.circle(I, (int(px), int(py)), 3, (0, 255, 0), thick)

# Class 2
for i in range(NTRAINING_SAMPLES, 2*NTRAINING_SAMPLES):
    px = trainData[i,0]
    py = trainData[i,1]
    cv.circle(I, (int(px), int(py)), 3, (255, 0, 0), thick)
```

- **Support vectors**

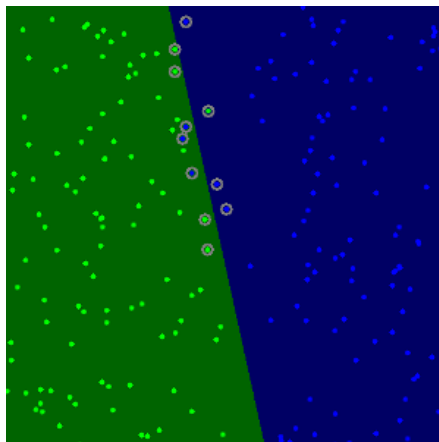
We use here a couple of methods to obtain information about the support vectors. The method `cv::ml::SVM::getSupportVectors` obtain all support vectors. We have used this methods here to find the training examples that are support vectors and highlight them.

```
thick = 2
sv = svm.getUncompressedSupportVectors()

for i in range(sv.shape[0]):
    cv.circle(I, (int(sv[i,0]), int(sv[i,1])), 6, (128, 128, 128), thick)
```

Results

- The code opens an image and shows the training examples of both classes. The points of one class are represented with light green and light blue ones are used for the other class.
- The SVM is trained and used to classify all the pixels of the image. This results in a division of the image in a blue region and a green region. The boundary between both regions is the separating hyperplane. Since the training data is non-linearly separable, it can be seen that some of the examples of both classes are misclassified; some green points lay on the blue region and some blue points lay on the green one.
- Finally the support vectors are shown using gray rings around the training examples.



You may observe a runtime instance of this on the [YouTube here](#).