

# Linear Hard Margin SVM for Classification and Pattern Recognition

André Ambrósio Boechat

Departamento de Automação e Sistemas  
Universidade Federal de Santa Catarina

Florianópolis, October 2012

# Contents

Introduction

Linear SVM Formulation

Problem Formulation

Lagrangian Formulation

Code

Conclusions



# Data-Driven Models

- Regression, pattern recognition, **classification**

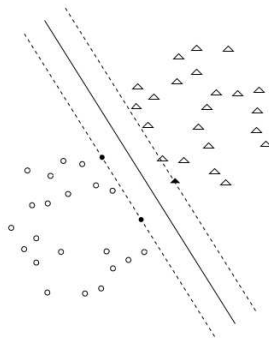
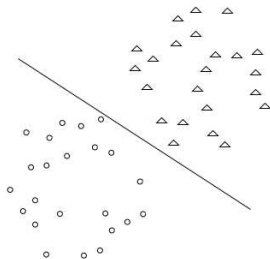
## Classification

Separation of samples in different **classes**, trying to find a possible **decision boundary**.

- Well-known classifiers
  - Naive Bayes classifier
  - Logistic regression
  - Artificial Neural networks (ANN)
  - Support Vector Machines (SVM)

# Support Vector Machines

- The foundations were recently developed (Vapnik, 1995)
- Great ability to generalize [Gunn, 1998]
- Many possible formulations
  - linear
  - nonlinear
  - regression
- **Large margin** classifier





## Geometrical Interpretation [Burges, 1998]

Training data composed of  $m$  samples

$$\{\mathbf{x}^{(i)}, y^{(i)}\}, \quad i = 1, \dots, m$$

$$\mathbf{x}^{(i)} \in \mathbb{R}^n$$

$$y^{(i)} \in \{-1, 1\}$$

We want to find a  $f(\mathbf{x})$  which determines a separating hyperplane

$$\mathbf{w}^T \mathbf{x} + b = 0 \tag{1}$$

For the linearly separable case, a hyperplane with the largest margin must obey the following constraints

$$\begin{aligned} \mathbf{w}^T \mathbf{x}^{(i)} + b &\geq 1, & \text{for } y^{(i)} &= 1 \\ \mathbf{w}^T \mathbf{x}^{(i)} + b &\leq -1, & \text{for } y^{(i)} &= -1 \end{aligned} \tag{2}$$



# Hyperplane with the Largest Margin

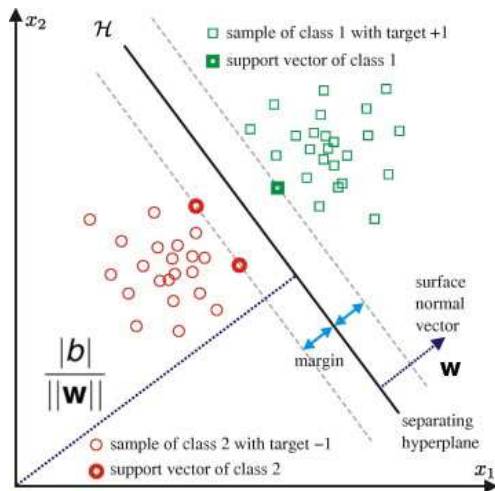


Figure: Adapted from [Fisch et al., 2010]



The support hyperplanes given by the constraints of Eq. (2) are

$$\begin{aligned}\mathcal{H}_1 : \quad \mathbf{w}^T \mathbf{x}^{(i)} + b &= 1 \\ \mathcal{H}_2 : \quad \mathbf{w}^T \mathbf{x}^{(i)} + b &= -1\end{aligned}\tag{3}$$

With simple geometry we could find that the sum of  $\mathcal{H}_1$  and  $\mathcal{H}_2$  margins is given by

$$\frac{2}{\|\mathbf{w}\|}$$

Rewritten the constraints of Eq. (2), we have

### SVM problem — Hard margin

$$\begin{aligned}\text{Minimize} \quad & \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & -y^{(i)} (\mathbf{w}^T \mathbf{x}^{(i)} + b) + 1 \leq 0, \quad \forall i\end{aligned}\tag{4}$$



# Lagrangian Formulation

## Primal problem

$$\begin{aligned} \max_{\lambda} \min_{\mathbf{w}, b} \quad & \mathcal{L}(\mathbf{w}, b, \lambda) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m \lambda_i y^{(i)} (\mathbf{w}^T \mathbf{x}^{(i)} + b) + \sum_{i=1}^m \lambda_i \\ & \lambda_i \geq 0 \end{aligned} \tag{5}$$

## Dual problem

$$\begin{aligned} \max_{\lambda} \quad & g(\lambda) = \inf_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \lambda) \\ & \lambda_i \geq 0 \end{aligned} \tag{6}$$





## Lagrangian Dual Problem

To find  $\inf_{\mathbf{w}, b} \mathcal{L}(\mathbf{w}, b, \lambda)$  the gradient with respect to  $\mathbf{w}$  and  $b$  must vanish, that gives the conditions

$$\mathbf{w} = \sum_{i=1}^m \lambda_i y^{(i)} \mathbf{x}^{(i)} \quad (7)$$

$$\sum_{i=1}^m \lambda_i y^{(i)} = 0 \quad (8)$$

### Dual problem

$$\begin{aligned} \max_{\lambda} \quad g(\lambda) &= \sum_{i=1}^m \lambda_i - \frac{1}{2} \sum_{i,j=1}^m \lambda_i \lambda_j y^{(i)} y^{(j)} \langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \rangle \\ \sum_{i=1}^m \lambda_i y^{(i)} &= 0 \\ \lambda_i &\geq 0 \end{aligned} \quad (9)$$



Considering the solution [Gunn, 1998]

$$\mathbf{w}^* = \sum_{i=1}^m \lambda_i y^{(i)} \mathbf{x}^{(i)}$$
$$b^* = -\frac{1}{2} \langle \mathbf{w}^*, \mathbf{x}^{(s_1)} + \mathbf{x}^{(s_2)} \rangle$$

- Those training samples for which  $\lambda_i > 0$  are called **support vectors** and lie on  $\mathcal{H}_1$  or  $\mathcal{H}_2$ .
- All other training samples have  $\lambda_i = 0$  and lie either on  $\mathcal{H}_1$  or  $\mathcal{H}_2$ , or on the half space determined by  $\mathcal{H}_1$  or  $\mathcal{H}_2$ .
- The support vectors are the **critical elements** of the training set.



## Karush-Kuhn-Tucker Conditions

Solving the SVM problem is equivalent to finding a solution to the KKT conditions. [Burges, 1998]

$$\begin{aligned}f_i(\mathbf{w}) &= -y^{(i)}(\mathbf{w}^T \mathbf{x}^{(i)} + b) + 1 \leq 0 \\ \lambda_i &\geq 0 \\ \lambda_i f_i(\mathbf{w}) &= \lambda_i \left[ -y^{(i)}(\mathbf{w}^T \mathbf{x}^{(i)} + b) + 1 \right] = 0 \\ \frac{\partial}{\partial \mathbf{w}} \mathcal{L} &= \mathbf{w} - \sum_{i=1}^m \lambda_i y^{(i)} \mathbf{x}^{(i)} = 0 \\ \frac{\partial}{\partial b} \mathcal{L} &= \sum_{i=1}^m \lambda_i y^{(i)} = 0\end{aligned} \tag{10}$$

Equation (10) is used to determine the  $b$  value.

*"If  $\tilde{\mathbf{w}}$ ,  $\tilde{b}$ ,  $\tilde{\lambda}$  satisfy KKT for a convex problem, then they are optimal"*

# Tools

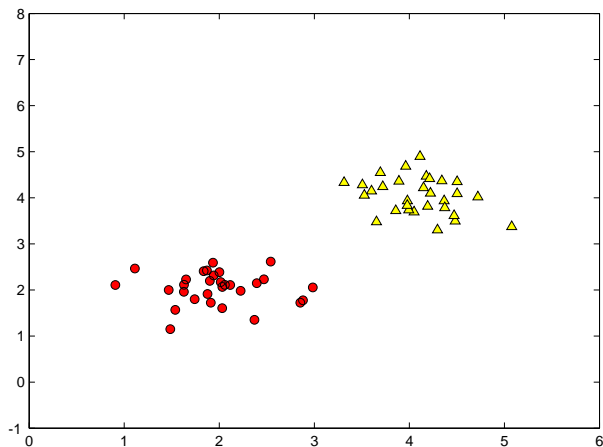
- SVM solvers/packages
  - LIBSVM
  - SVM light
  - Matlab SVM toolbox
  - List of softwares:  
[www.support-vector-machines.org/SVM\\_soft.html](http://www.support-vector-machines.org/SVM_soft.html)
- **CVX**, a convex modeling framework for Matlab
  - problem description very similar to the mathematical formulation

# Dataset Example

```
%% Linear separable data samples generation for training.  
% Features dimension  
n = 2;  
% Number of samples  
m = 2*30;  
% Center of the classes  
c1 = [2 2];  
c2 = [4 4];  
% Standard deviation from center  
stdc = [.4 .4];  
% Data samples -> X is MxN  
X1 = repmat(c1, m/2, 1) + repmat(stdc, m/2, 1) .* randn(m/2, n);  
X2 = repmat(c2, m/2, 1) + repmat(stdc, m/2, 1) .* randn(m/2, n);  
X = [X1; X2];  
% Labels -> Y is Mx1  
Y = [ones(m/2, 1); -1*ones(m/2, 1)];
```

○○○  
○○○

# Linear Separable Dataset



# SVM Primal Problem

```
function [w, b] = svm_primal(X, Y)

[m, n] = size(X);
%% SVM formulation
cvx_begin
    variables w(1,n) b(1)
    minimize(pow_pos(norm(w, 2), 2))
    - Y .* (X * w' + ones(m,1)*b) + ones(m,1) <= zeros(m,1);
cvx_end
```

# SVM Dual Problem

```

function [w, b] = svm_lagrangian(X, Y)

[m, n] = size(X);
%% Dual problem of the Lagrangian formulation.
Z = repmat(Y, 1, n) .* X;
H = Z * Z'; % MxM matrix
cvx_begin
    variables lambda(m, 1)
    %maximize ( sum(lambda) - .5 * lambda' * H * lambda )
    maximize ( sum(lambda) - .5 * quad_form(lambda, H) )
    lambda'*Y == 0;
    lambda >= zeros(m, 1);
cvx_end

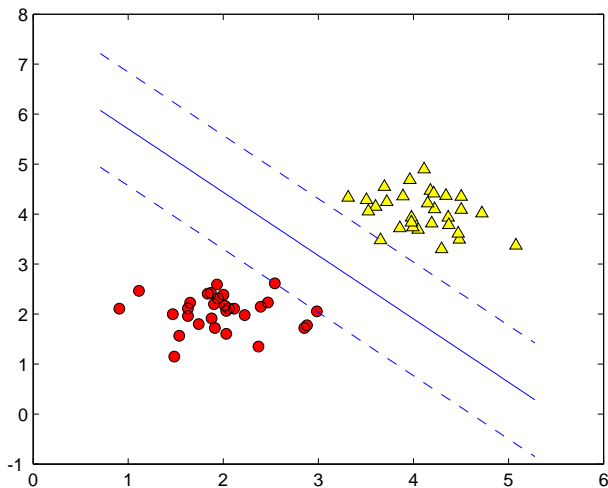
% w is 1xN
w = (lambda .* Y)' * X;
% Non-zero lagrangian multipliers.
l1 = find(lambda > 1e-6 & Y == 1);
l2 = find(lambda > 1e-6 & Y == -1);
% b calculation as Gunn1998 suggested.
b = - .5 * w * (X(l1(1), :) + X(l2(1), :))';

```



ooo  
oooo

# Problem Solution



ooo  
oooo

---

```
number of iterations    = 22
primal objective value  = -2.01515477e+00
dual  objective value  = -2.01515479e+00
gap := trace(XZ)       = 1.58e-08
relative gap           = 3.13e-09
actual relative gap    = 3.13e-09
rel. primal infeas     = 2.76e-12
rel. dual  infeas     = 1.00e-12
Total CPU time (secs)  = 0.61
CPU time per iteration = 0.03
termination code       = 0
```

---

```
number of iterations    = 14
primal objective value  = -1.00757737e+00
dual  objective value  = -1.00757740e+00
gap := trace(XZ)       = 3.09e-08
relative gap           = 1.02e-08
actual relative gap    = 1.02e-08
rel. primal infeas     = 1.56e-12
rel. dual  infeas     = 1.01e-12
Total CPU time (secs)  = 0.20
CPU time per iteration = 0.01
termination code       = 0
```

---



# Conclusions

- SVM is a famous technique for classification problems
- The SVM problem consists in looking for separating hyperplane with the largest margins
- It leads to a convex problem
- There are advantages in using the Lagrangian formulation
- The solution of the dual problem is optimal (KKT conditions)
- With a little modification, the presented formulation can handle more complex problems

## Further Reading



Burges, C. (1998).

A tutorial on support vector machines for pattern recognition.

*Data mining and knowledge discovery*, 43:1–43.



Fisch, D., Kühbeck, B., Sick, B., and Ovaska, S. J. (2010).

So near and yet so far: New insight into properties of some well-known classifier paradigms.

*Information Sciences*, 180(18):3381 – 3401.



Gunn, S. (1998).

Support vector machines for classification and regression.  
*Technical Report 2*.