Soft SVM

Dr. Fayyaz ul Amir Afsar Minhas
PIEAS Biomedical Informatics Research Lab
Department of Computer and Information Sciences
Pakistan Institute of Engineering & Applied Sciences
PO Nilore, Islamabad, Pakistan
http://faculty.pieas.edu.pk/fayyaz/

Matrix formulation of SVM Problem

$$\max_{\alpha_i \geq 0} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^T} \mathbf{x}^{(j)} \right\}$$

$$s.t \qquad \alpha_i \geq 0$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

$$max_{\boldsymbol{\alpha}}\ \mathbf{1}^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\alpha}$$

Subject to:

$$\boldsymbol{\alpha} \succcurlyeq \mathbf{0}$$

$$\boldsymbol{y}^T \boldsymbol{\alpha} = 0$$

- If we define the following:
  - $\boldsymbol{\alpha} = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_N]^{\boldsymbol{T}}$
  - $\boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_N]^{\boldsymbol{T}}$
  - $\mathbf{1}_N = [1 \quad 1 \quad \ldots \quad 1]^{\boldsymbol{T}}$
  - $\boldsymbol{X}_{(d \times N)} = [\boldsymbol{x}_{\mathbf{1}} y_1 \quad \boldsymbol{x}_{\mathbf{2}} y_2 \quad \cdots \quad \boldsymbol{x}_N y_N]$

Solving the SVM using QP

- CVXOPT is a Python package that implements a quadratic programming solver
  - http://abel.ee.ucla.edu/cvxopt/userguide/coneprog.html#quadratic-programming
- cvxopt.solvers.qp(*P*, *q*[, *R*, *s*[, *U*, *v*[, *solver*[, *initvals*]]]])
  - Solves the following problem for $\boldsymbol{z}$:

$$min_{\boldsymbol{z}} \frac{1}{2} \boldsymbol{z}^T \boldsymbol{P} \boldsymbol{z} + \boldsymbol{q}^T \boldsymbol{z}$$

Subject to:

$$\boldsymbol{R}\boldsymbol{z} \preccurlyeq \boldsymbol{s}$$
$$\boldsymbol{U}\boldsymbol{z} = \boldsymbol{v}$$

$\leftarrow$

$\boldsymbol{z} = \boldsymbol{\alpha}$
$\boldsymbol{P} = \boldsymbol{X}^T \boldsymbol{X}$
$\boldsymbol{q} = -\boldsymbol{1}_N$
$\boldsymbol{R} = -\boldsymbol{I}_{N \times N}$
$\boldsymbol{s} = \boldsymbol{0}_N$
$\boldsymbol{U} = \boldsymbol{y}^T$
$\boldsymbol{v} = 0$

SVM Problem

$$max_{\boldsymbol{\alpha}} \boldsymbol{1}^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{\alpha}$$

Subject to:

$$\boldsymbol{\alpha} \succcurlyeq \boldsymbol{0}$$
$$\boldsymbol{y}^T \boldsymbol{\alpha} = 0$$

Programming the SVM

Example: Solution of the OR problem

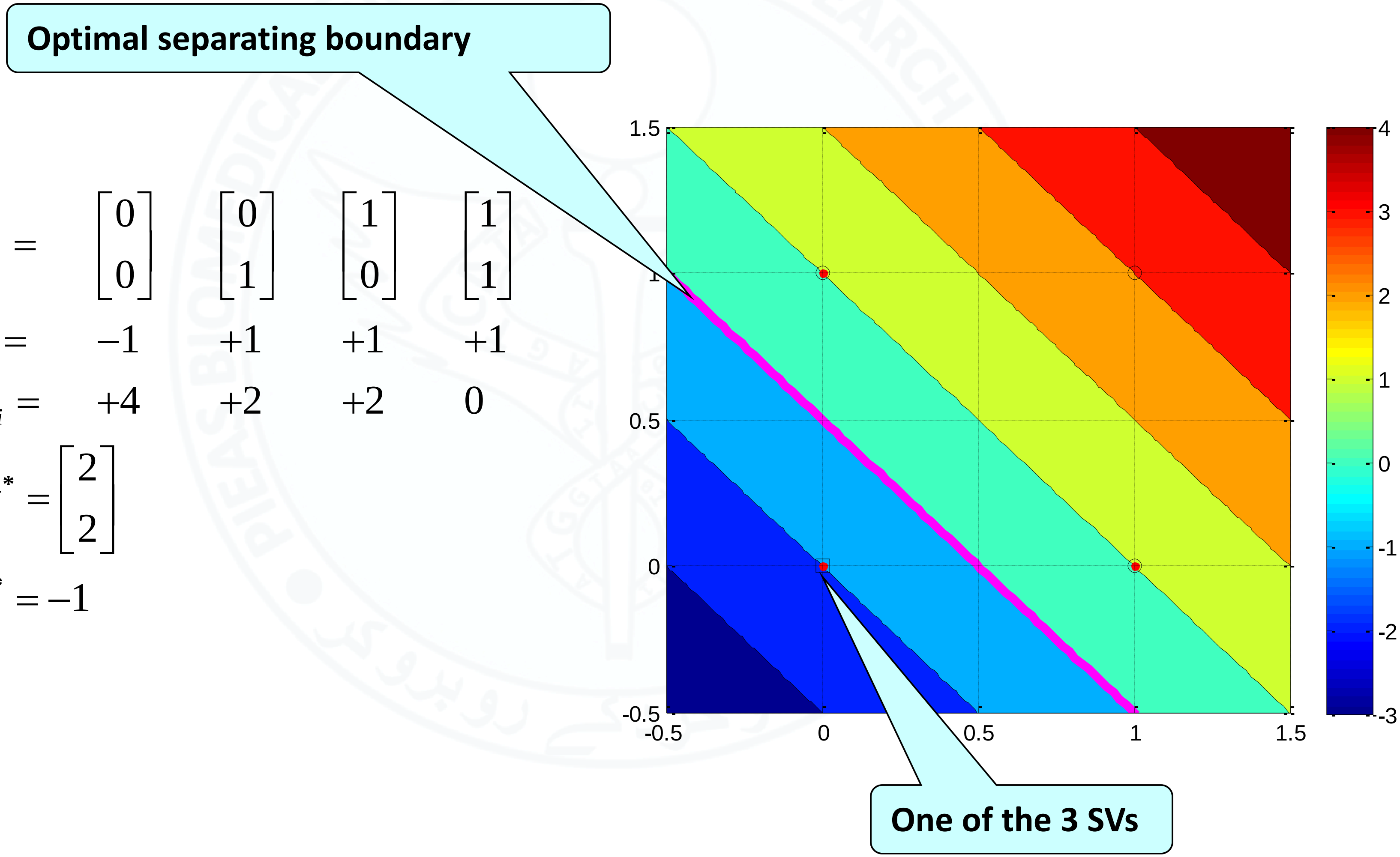
$$\mathbf{x}^i = \begin{bmatrix}0\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\end{bmatrix} \quad \begin{bmatrix}1\\0\end{bmatrix} \quad \begin{bmatrix}1\\1\end{bmatrix}$$

$$y = \quad -1 \quad +1 \quad +1 \quad +1$$

$$\alpha_i = \quad +4 \quad +2 \quad +2 \quad 0$$

$$\mathbf{w}^* = \begin{bmatrix}2\\2\end{bmatrix}$$

$$b^* = -1$$



Handling Non-Separable Patterns

- All the derivation till now was based on the requirement that the data be linearly separable
  - Practical Problems are Non-Separable
- Non-separable data can be handled by relaxing the constraint

$$y_i\left(\mathbf{w}^T\mathbf{x}+b\right)\geq 1$$

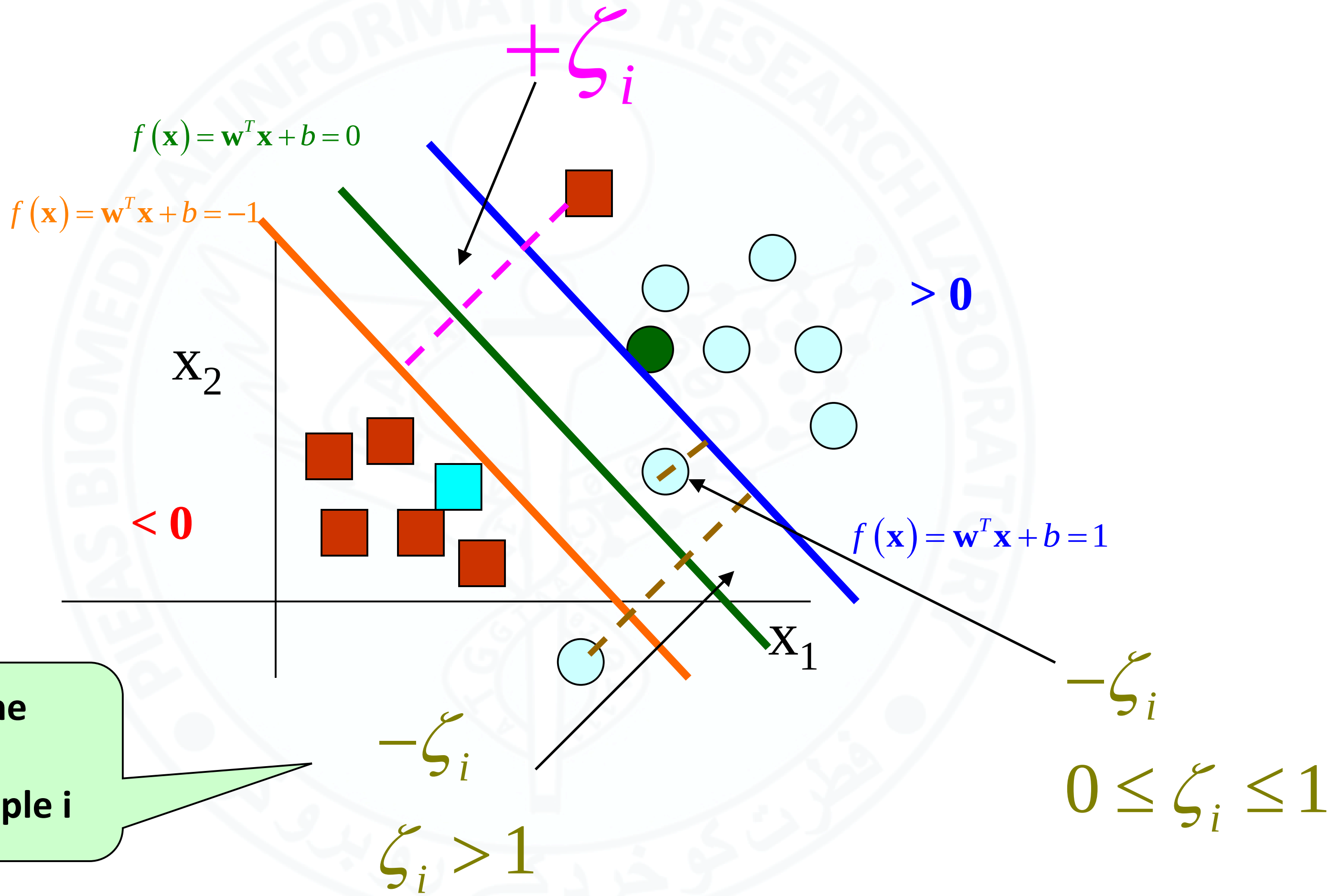
- This is achieved as

$$\left.\begin{aligned}\mathbf{w}^T\mathbf{x}^{(i)}+b &\geq 1-\zeta_i \qquad \forall y_i=+1\\ \mathbf{w}^T\mathbf{x}^{(i)}+b &\geq -1+\zeta_i \qquad \forall y_i=-1\end{aligned}\right\} \quad y_i\left(\mathbf{w}^T\mathbf{x}^{(i)}+b\right)\geq 1-\zeta_i$$

$$\zeta_i \geq 0$$

Slack Variables

Handling Non-Separable Patterns (Soft Margin)



ζ_i tells us about the extent of margin violation of example i

Handling Non-Separable Patterns (Soft Margin)

- Our objective in designing a SVM here is to maximize the margin while minimizing the misclassification error
- The misclassification error can be written as:

$$\Phi(\zeta) = \sum_{i=1}^{N} I(\zeta_i - 1) \qquad I(\zeta) = \begin{cases} 0 & \zeta \leq 0 \\ 1 & else \end{cases}$$

- Since this function is non-linear and non-convex, therefore we choose to use an approximate given by

$$\Phi(\zeta) = \sum_{i=1}^{N} \zeta_i$$

Soft Margin SVM as an Optimization Problem

- The overall optimization problem can be written as

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i$$

$$s.t. \quad y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \geq 1 - \zeta_i$$

$$\zeta_i \geq 0$$

Weight of the penalty due to margin violation

Or

$$\min_{\mathbf{w},b} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i$$

$$s.t. \quad 1 - \zeta_i - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \leq 0$$

$$-\zeta_i \leq 0$$

The objective function now optimizes two conflicting objectives: Maximization of the margin and minimization of the margin violations

Soft Margin SVM as an Optimization Problem

- Writing the constraints as losses, we have

$$\min_{\mathbf{w},b,\zeta} \quad M = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i + \sum_{i=1}^{N}L_i + \sum_{i=1}^{N}Z_i$$

$$L_i = \max_{\alpha_i} \quad \alpha_i\left\{1 - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\} = \begin{cases} 0 & 1 - \zeta_i - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right) \le 0 \\ \infty & else \end{cases}, \qquad \alpha_i \ge 0$$

$$Z_i = \max_{\beta_i} \quad \beta_i\left\{-\zeta_i\right\} = \begin{cases} 0 & -\zeta_i \le 0 \\ \infty & else \end{cases}, \qquad \beta_i \ge 0$$

Thus

$$\min_{\mathbf{w},b,\zeta}\max_{\alpha_i,\beta_i} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i + \sum_{i=1}^{N}\alpha_i\left\{1 - \zeta_i - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)} + b\right)\right\} - \sum_{i=1}^{N}\beta_i\zeta_i$$

$$s.t \quad \alpha_i \ge 0, \beta_i \ge 0$$

- This is the primal form of the optimization problem

Soft Margin SVM as an Optimization Problem

- The Dual Form is

$$\max_{\alpha_i,\beta_i} \min_{\mathbf{w},b,\zeta} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i + \sum_{i=1}^{N}\alpha_i\left\{1-\zeta_i - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)}+b\right)\right\} - \sum_{i=1}^{N}\beta_i\zeta_i$$

$$s.t \quad \alpha_i \geq 0, \beta_i \geq 0$$

- Since the optimization problem is convex, therefore the optimal values of the dual and primal are equal
- Therefore the optimization problem can be written solved in its dual form

Soft Margin SVM as an Optimization Problem

$$\max_{\alpha_i,\beta_i}\min_{\mathbf{w},b,\zeta} \quad \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i + \sum_{i=1}^{N}\alpha_i\left\{1-\zeta_i - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)}+b\right)\right\} - \sum_{i=1}^{N}\beta_i\zeta_i$$

$$= \max_{\alpha_i,\beta_i} \theta_D\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)$$

$$\theta_D\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right) = \min_{\mathbf{w},b,\zeta}\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\zeta_i + \sum_{i=1}^{N}\alpha_i\left\{1-\zeta_i - y_i\left(\mathbf{w}^T\mathbf{x}^{(i)}+b\right)\right\} - \sum_{i=1}^{N}\beta_i\zeta_i$$

Solving

$$\frac{\partial\theta_D}{\partial\mathbf{w}} = \mathbf{w} - \sum_{i=1}^{N}\alpha_i y_i \mathbf{x}^{(i)} = 0 \qquad \Rightarrow \mathbf{w} = \sum_{i=1}^{N}\alpha_i y_i \mathbf{x}^{(i)}$$

$$\frac{\partial\theta_D}{\partial b} = \sum_{i=1}^{N}\alpha_i y_i = 0 \qquad \Rightarrow \sum_{i=1}^{N}\alpha_i y_i = 0$$

$$\frac{\partial\theta_D}{\partial\zeta_i} = C - \alpha_i - \beta_i = 0 \qquad \Rightarrow \alpha_i + \beta_i = C$$

Soft Margin SVM as an Optimization Problem

$$\max_{\alpha_i,\beta_i \geq 0} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^T} \mathbf{x}^{(j)} \right\}$$

$$s.t \qquad \alpha_i \geq 0, \beta_i \geq 0$$

$$\alpha_i + \beta_i = C, \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

Simplifying further

$$\max_{\alpha_i,\beta_i \geq 0} \left\{ \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}^{(i)^T} \mathbf{x}^{(j)} \right\}$$

$$s.t \qquad 0 \leq \alpha_i \leq C, \qquad \sum_{i=1}^{N} \alpha_i y_i = 0$$

- This problem can be solved using standard optimization packages

Soft Margin SVM as an Optimization Problem

- Some Observations
  - α_i will be non-zero (positive) only for the points that are support vectors
  - $\beta_i = C - \alpha_i$ will be zero for the points that violate the margin condition
  - C is the upper bound on α_i
  - C is the weight of the penalty of the term representing margin violation
    - If C is small, then more margin violations will occur
    - If C is large, lesser margin violations will result
  - C can be found out through cross-validation

Example

$$\mathbf{x}^i = \begin{bmatrix}0\\0\end{bmatrix} \quad \begin{bmatrix}0\\1\end{bmatrix} \quad \begin{bmatrix}1\\0\end{bmatrix} \quad \begin{bmatrix}1\\1\end{bmatrix} \quad \begin{bmatrix}0.8\\0.5\end{bmatrix}$$

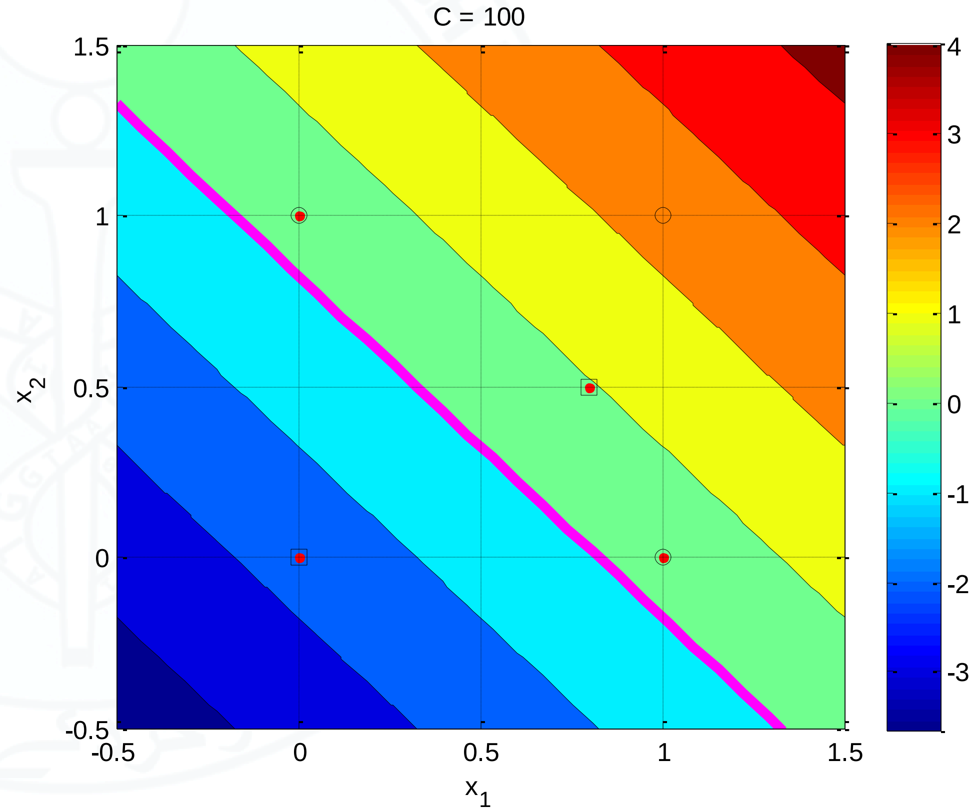
$$C = 100$$

$$y = \quad -1 \quad +1 \quad +1 \quad +1 \quad -1$$

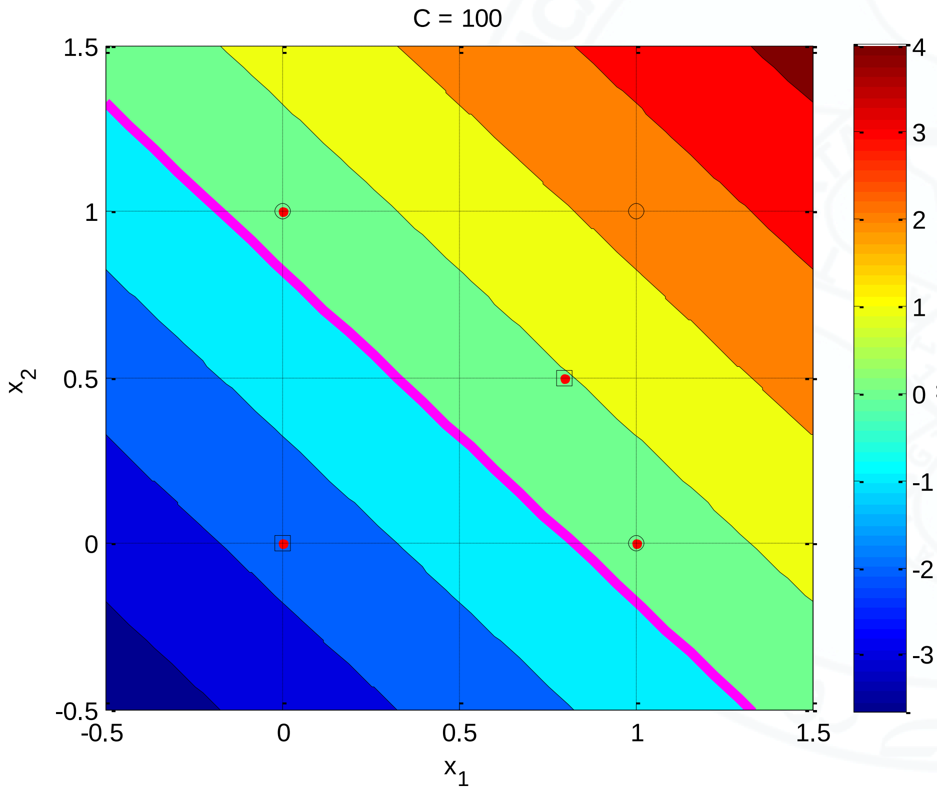
$$\alpha_i = \quad +34 \quad +52 \quad +82 \quad 0 \quad +100$$

$$\mathbf{w}^* = \begin{bmatrix}2\\2\end{bmatrix}$$

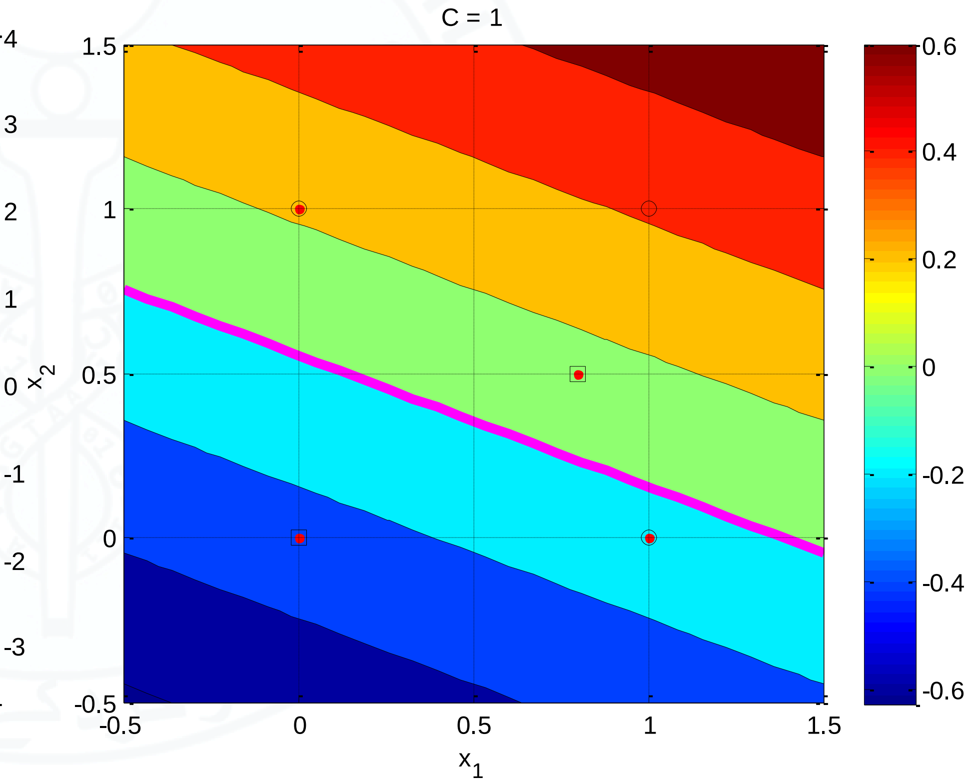
$$b^* = -1.65$$



Example...



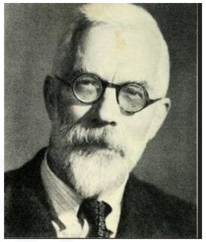
C = 100
1.5
1
0.5
0
-0.5
-0.5
0
0.5
1
1.5
x_1
x_2
4
3
2
1
0
-1
-2
-3



C = 1
1.5
1
0.5
0
-0.5
-0.5
0
0.5
1
1.5
x_1
x_2
0.6
0.4
0.2
0
-0.2
-0.4
-0.6

SVMs uptil now

- Vapnik and Chervonenkis:
  - Hard SVM (1962)
  - Theoretical foundations for SVMs
- Corinna Cortes
  - Soft SVM (1995)
- Enter: Bernard Scholkopf (1997)
  - Complete Kernel trick!
  - Kernels not only allow nonlinear boundaries but also allow representation of non-vectoral data


R. A. Fisher
1890-1962


Rosenblatt
1928-1971


V. Vapnik
1936 -


Chervonenkis
1938 - 2014


C. Cortes
1961 -


Scholkopf
1968 -

http://www.svms.org/history.html

Reading

- Sections 10.1-10.3
- Sections 13.1-13.3
- Alpaydin, Ethem. Introduction to Machine Learning. Cambridge, Mass. MIT Press, 2010.

We want to make a machine that will be proud of us.

- Danny Hillis