UNIVERSITY OF LEEDS

Class: Machine Learning

Support Vector Machines – part 2

Instructor: Matteo Leonetti

Learning outcomes

UNIVERSITY OF LEEDS

- Define Soft-Margin SVMs
- Project a given dataset to a higher-dimensional space

The SVM Formulation

Margin as large as possible

minimise: $\frac{1}{2}\|\boldsymbol{w}\|^2$

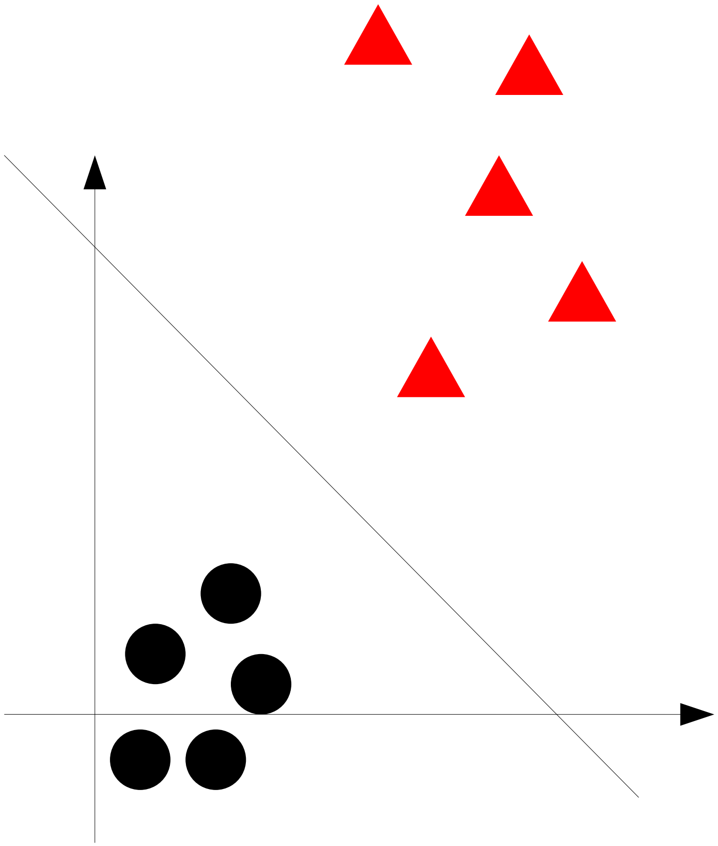
Subject to the constraints: $t_i(\boldsymbol{w}^T \boldsymbol{x}_i + w_0) \geq 1$

Every point is on the correct side, no point is on the hyperplane

Why SUPPORT VECTOR Machine?

UNIVERSITY OF LEEDS

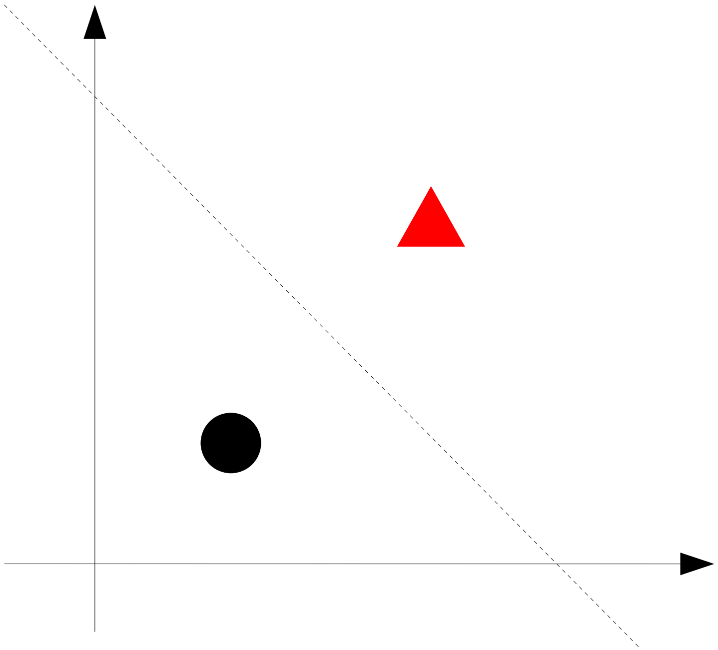
Consider this dataset:



Why SUPPORT VECTOR Machine?

UNIVERSITY OF LEEDS

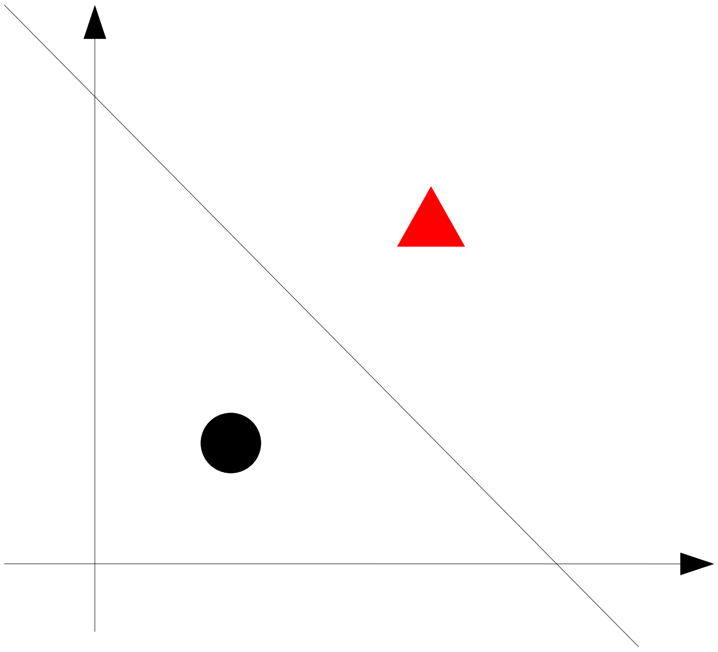
Consider this dataset:



Does the optimal separating line change if I remove all but the closest points?

Why SUPPORT VECTOR Machine?

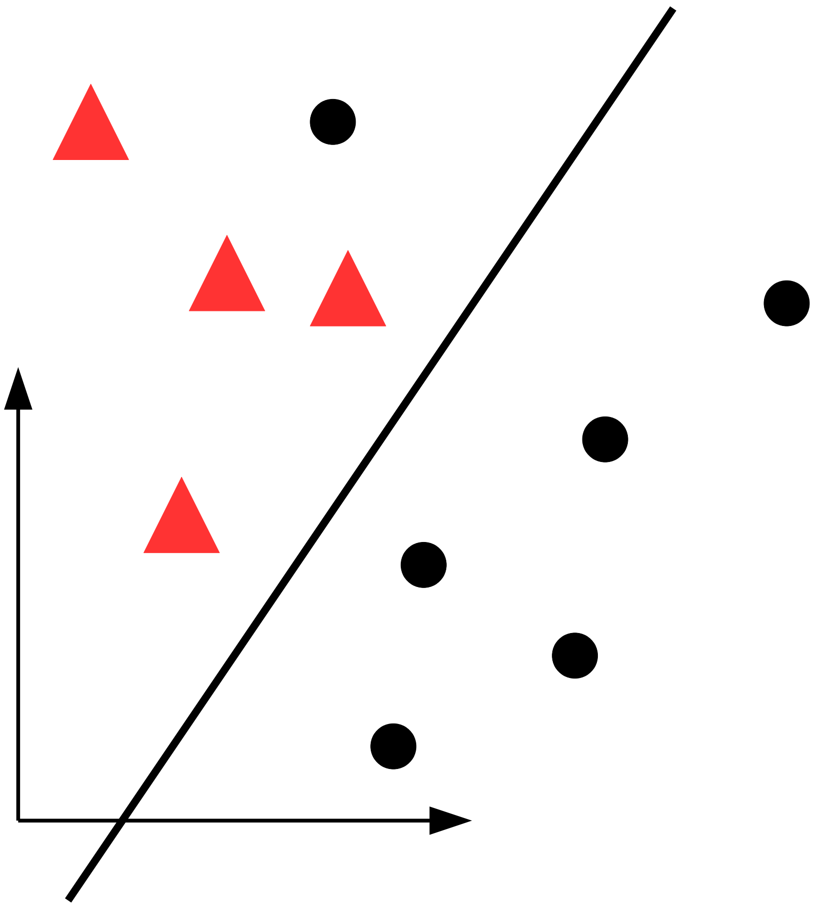
Consider this dataset:



No!

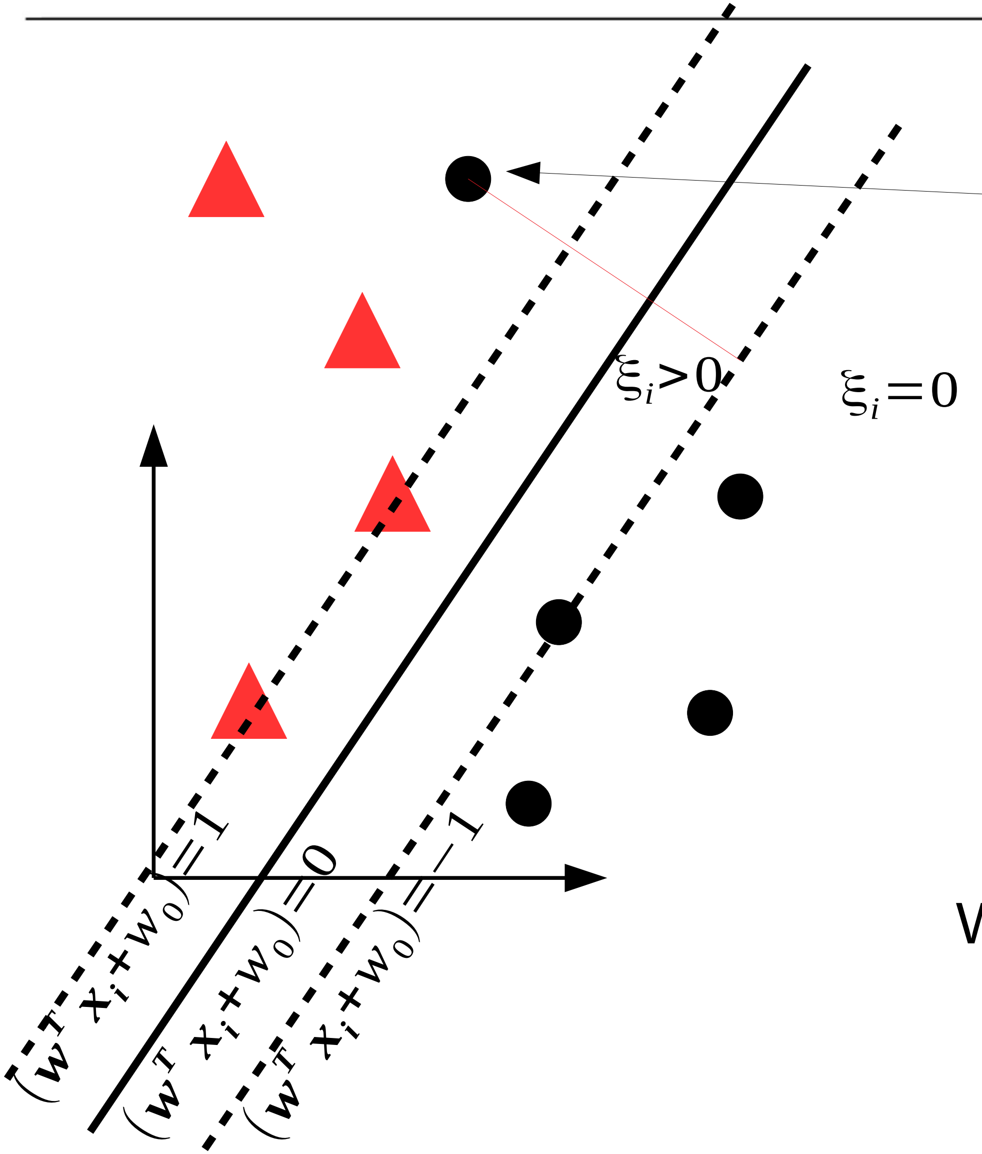
Back to linear separability

UNIVERSITY OF LEEDS



Is this classifier acceptable?

Slack Variables



For this point we want to allow the corresponding constraint to be *softened*:

$$t_i(\boldsymbol{w}^T \boldsymbol{x_i} + w_0) \geq 1$$

One way to do this is to introduce an additional, positive, variable:

$$t_i(\boldsymbol{w}^T \boldsymbol{x_i} + w_0) \geq 1 - \xi_i$$

With the additional constraints:

$$\xi_i \geq 0$$

Soft margin

Min $\|\boldsymbol{w}\|^2 + C \sum_i \xi_i$

How much we are violating the constraints

Weight of violations

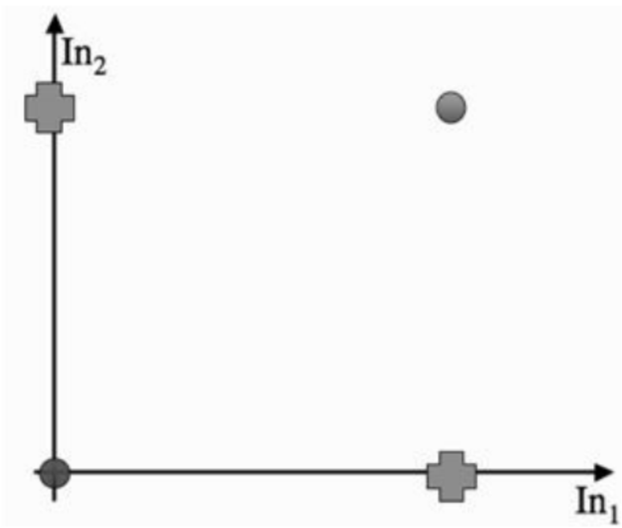
Subject to:

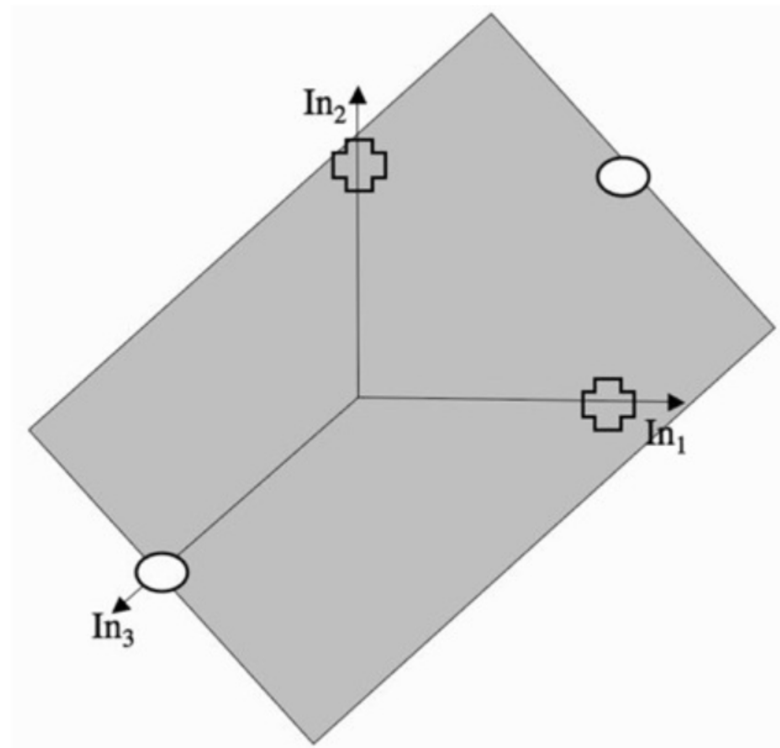
$$t_i(\boldsymbol{w}^T \boldsymbol{x_i} + w_0) \geq 1 - \xi_i \qquad \xi_i \geq 0$$

Sensible to outliers!

$C = \infty$ Hard margin

Linear separability… revisited





Wait, what?!? More dimensions seem to help!

Where do we get the extra dimensions from?

Example: Polynomial features

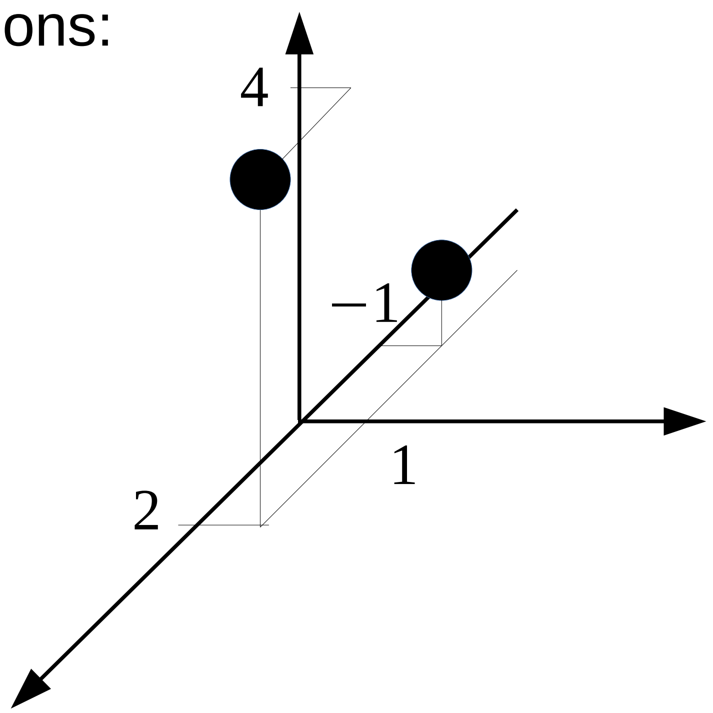
Let's take 2 points in 1 dimension:

$\langle -1 \rangle, \langle 2 \rangle$

and project them in 3 dimensions:

In general: $\langle 1, x, x^2 \rangle$

Our points: $\langle 1, -1, 1 \rangle, \langle 1, 2, 4 \rangle$



Example: Polynomial features

The original dataset has 1 variable:

$$\langle x_{1,} t_1 \rangle, \langle x_{2,} t_2 \rangle, \ldots, \langle x_N, t_N \rangle$$

But we want a higher dimensional space...

Let's use polynomial features: $\Phi_i(x) = x^i$

Our points become:

$$\langle 1, x_{1,} x_{1,}^2 x_{1,}^3 \ldots, x_{1,}^d t_1 \rangle, \langle 1, x_{2,} x_{2,}^2 x_{2,}^3 \ldots, x_{2,}^d t_2 \rangle, \ldots, \langle 1, x_N, x_N^2, x_N^3, \ldots, x_N^d, t_N \rangle$$

Features

How can we add extra dimensions?

Original point: x

Define a set of functions $\Phi_i(x)$

New point: $\Phi(x)=\langle\Phi_0(x),\Phi_1(x),\Phi_2(x),\Phi_3(x),\ldots,\Phi_n(x)\rangle$

Substitution

dataset:

$$\langle x_i, t_i \rangle = \langle -1, 1 \rangle, \langle 2, -1 \rangle$$

$$\langle \boldsymbol{x}_i, t_i \rangle = \langle 1, -1, 1, 1 \rangle, \langle 1, 2, 4, -1 \rangle$$

problem:

min $\frac{1}{2} \|\langle w_1 \rangle\|^2 = \frac{1}{2} w_1^2$

s.t.: $1 \cdot (-1 \cdot w_1 + w_0) \geq 1$

$-1 \cdot (2 \cdot w_1 + w_0) \geq 1$

min $\frac{1}{2} \|\langle w_{1,} w_{2,} w_3 \rangle\|^2$

s.t.: $1 \cdot \left(\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + w_0 \right) \geq 1$

$-1 \cdot \left(\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} + w_0 \right) \geq 1$

Substitution

minimise: $$\frac{1}{2}\|\boldsymbol{w}\|^2$$

Subject to the constraints: $$t_i\left(\boldsymbol{w}^{\boldsymbol{T}}\Phi\left(\boldsymbol{x}_i\right)+w_0\right)\geq 1$$

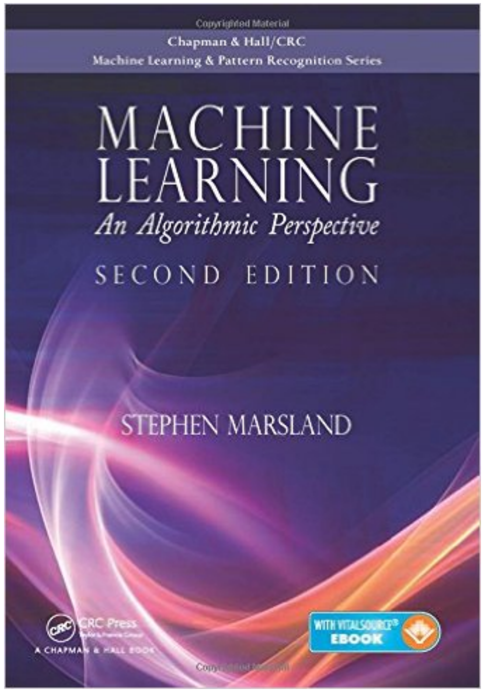
This way we would have a higher dimensional problem, which is also more difficult to solve.

Is there a better formulation?

Conclusion

Learning outcomes

- Define Soft-Margin SVMs
- Project a given dataset to a higher-dimensional space


Copyrighted Material
Chapman & Hall/CRC
Machine Learning & Pattern Recognition Series
MACHINE LEARNING
An Algorithmic Perspective
SECOND EDITION
STEPHEN MARSLAND
CRC Press
A CHAPMAN & HALL BOOK
WITH VITALSOURCE® EBOOK
Copyrighted Material

Chapter 8.2