UNIVERSITY OF LEEDS

# Class: Machine Learning

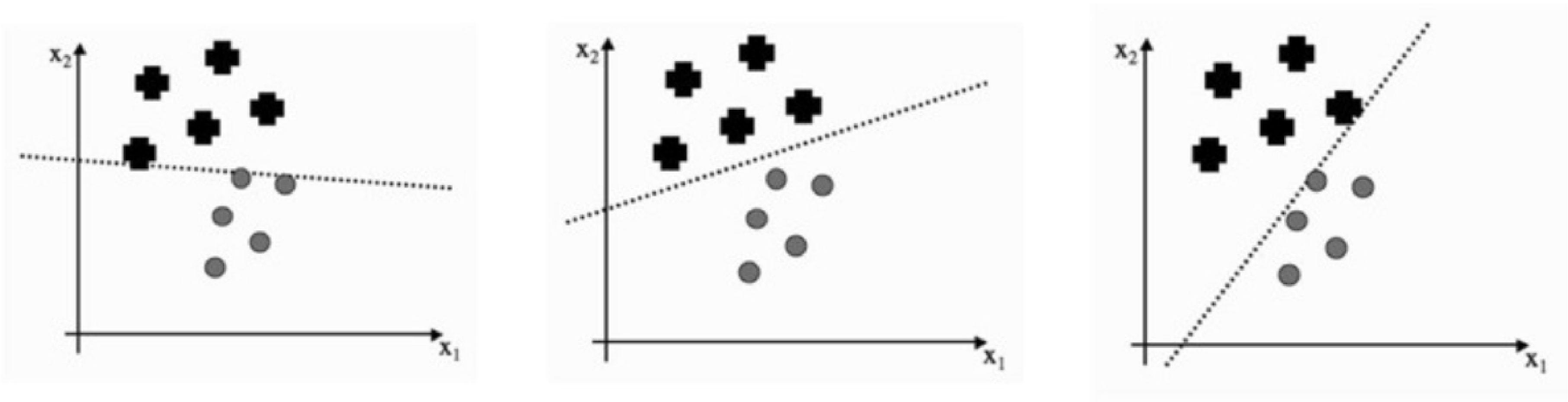
# Support Vector Machines


Instructor: Matteo Leonetti

# Learning outcomes

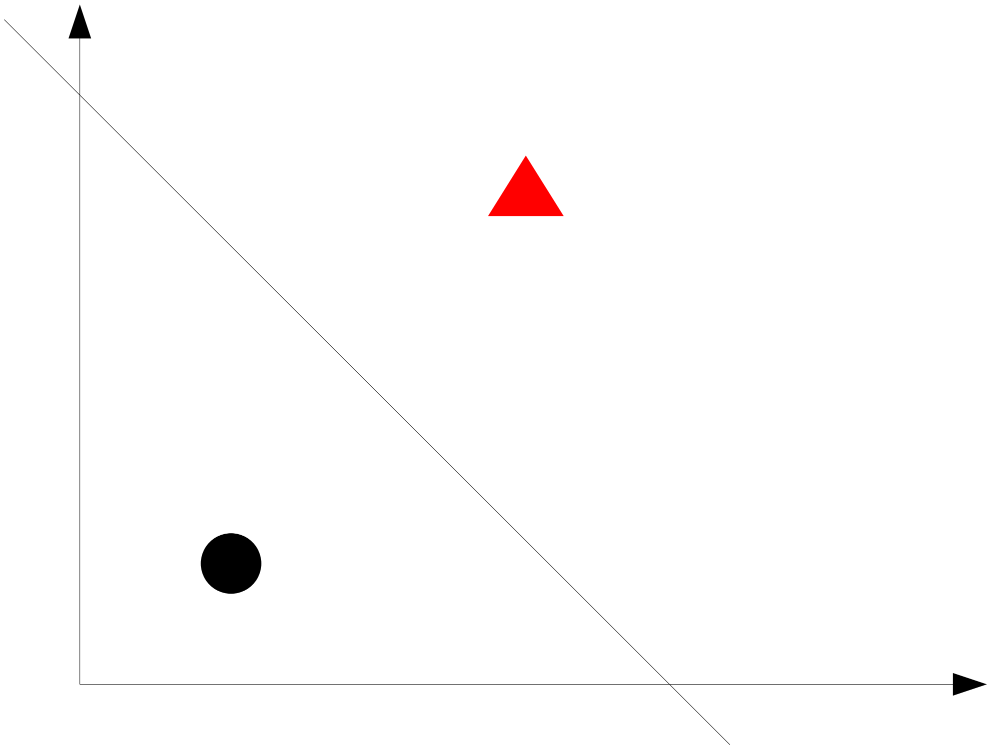
- Derive the formulation of support vector machines as a constrained optimisation problem

# Multiple separation boundaries




Is any one better than the others?

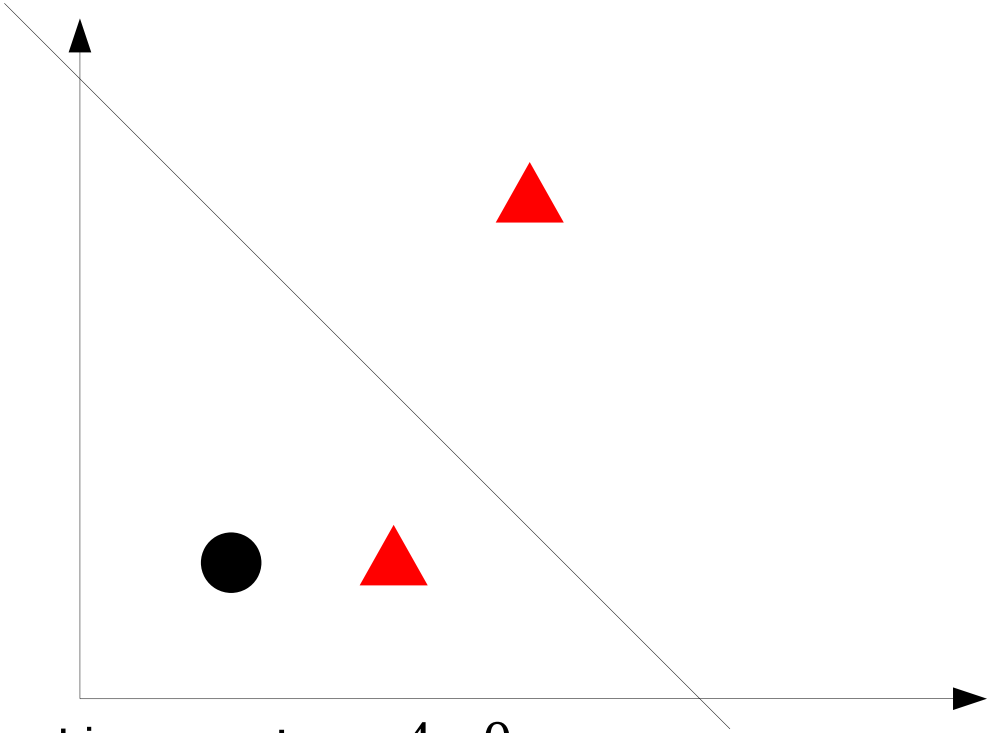
# The Best Discriminating Boundary



What is special about this line?

# The Constraints

UNIVERSITY OF LEEDS



Line: $x_1 + x_2 - 4 = 0$

Classes: {-1,1}

Let's redefine the output of the classifier:

$$y(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{w}^T \boldsymbol{x} + w_0 \geq 0 \\ -1 & \text{if } \boldsymbol{w}^T \boldsymbol{x} + w_0 < 0 \end{cases}$$

Example

point: $\langle 1,1 \rangle$ desired class: $t = -1$ output: $1+1-4 = -2 \leq 0 \Rightarrow y = -1$

point: $\langle 3,3 \rangle$ desired class: $t = 1$ output: $3+3-4 = 2 > 0 \Rightarrow y = 1$

point: $\langle 2,1 \rangle$ desired class: $t = 1$ output: $2+1-4 = -1 \leq 0 \Rightarrow y = -1$

# The Constraints



Line: $x_1 + x_2 - 4 = 0$

point: $\langle 1,1 \rangle$ desired class: $t = -1$ output: $1+1-4=-2 \leq 0 \Rightarrow y=-1$

point: $\langle 3,3 \rangle$ desired class: $t = 1$ output: $3+3-4=2>0 \Rightarrow y=1$

point: $\langle 2,1 \rangle$ desired class: $t = 1$ output: $2+1-4=-1 \leq 0 \Rightarrow y=-1$

When the point is classified correctly:

$$ty = 1$$

Since: $y(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{w}^T \boldsymbol{x} + w_0 \geq 0 \\ -1 & \text{if } \boldsymbol{w}^T \boldsymbol{x} + w_0 < 0 \end{cases}$ This is the same as:

$$t\left(\boldsymbol{w}^T \boldsymbol{x} + w_0\right) \geq 0$$

# Canonical form

UNIVERSITY OF LEEDS

$$t(\boldsymbol{w}^T \boldsymbol{x} + w_0) = 0$$

Means the classifier is undecided, and should be avoided.

We can do so by imposing that: $t(\boldsymbol{w}^T \boldsymbol{x} + w_0) \geq \epsilon$ with $\epsilon > 0$

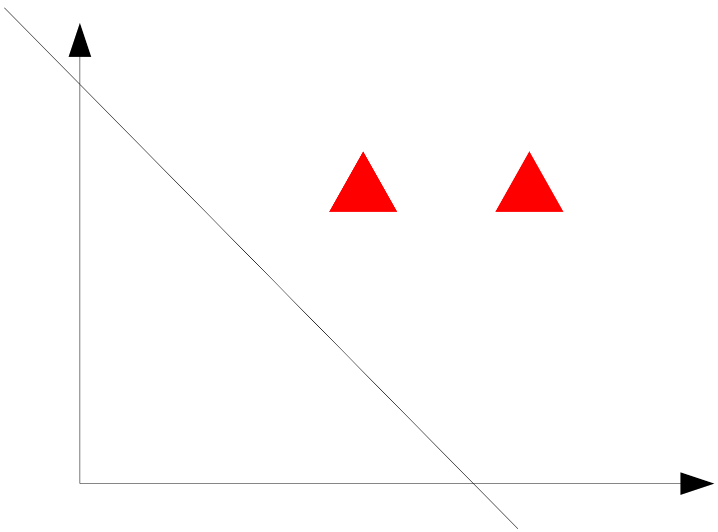
By dividing both sides of the inequality by a constant, we can make ε any number (other than 0). We like 1:

$$\boxed{t(\boldsymbol{w}^T \boldsymbol{x} + w_0) \geq 1}$$

This is called the *canonical* form of the constraints.

# Example of Constraint

UNIVERSITY OF LEEDS



Line: $x_1 + x_2 - 4 = 0$

$$\boldsymbol{w} = \langle -4, 1, 1 \rangle$$

The closest point is $\langle 3, 3 \rangle$

$$t(\boldsymbol{w}^T \langle 3,3 \rangle + w_0) = 1(\langle 1,1 \rangle^T \cdot \langle 3,3 \rangle - 4) = 2$$
$$t(\boldsymbol{w}^T \langle 4,3 \rangle + w_0) = 1(\langle 1,1 \rangle^T \cdot \langle 4,3 \rangle - 4) = 3$$

So right now, for any point x: $t(\boldsymbol{w}^T \boldsymbol{x} + w_0) \geq 2$

I can rescale the weights so that for the closest point: $t(\boldsymbol{w}^T \boldsymbol{x} + w_0) = 1$

$$\frac{t(\boldsymbol{w}^T \langle 3,3 \rangle + w_0)}{2} = 1\left(\frac{\langle 1,1 \rangle^T}{2} \cdot \langle 3,3 \rangle - \frac{4}{2}\right) = 1(\langle 0.5, 0.5 \rangle^T \langle 3,3 \rangle - 2) = 1$$

So, with the new vector: $\boldsymbol{w}' = \langle -2, 0.5, 0.5 \rangle$

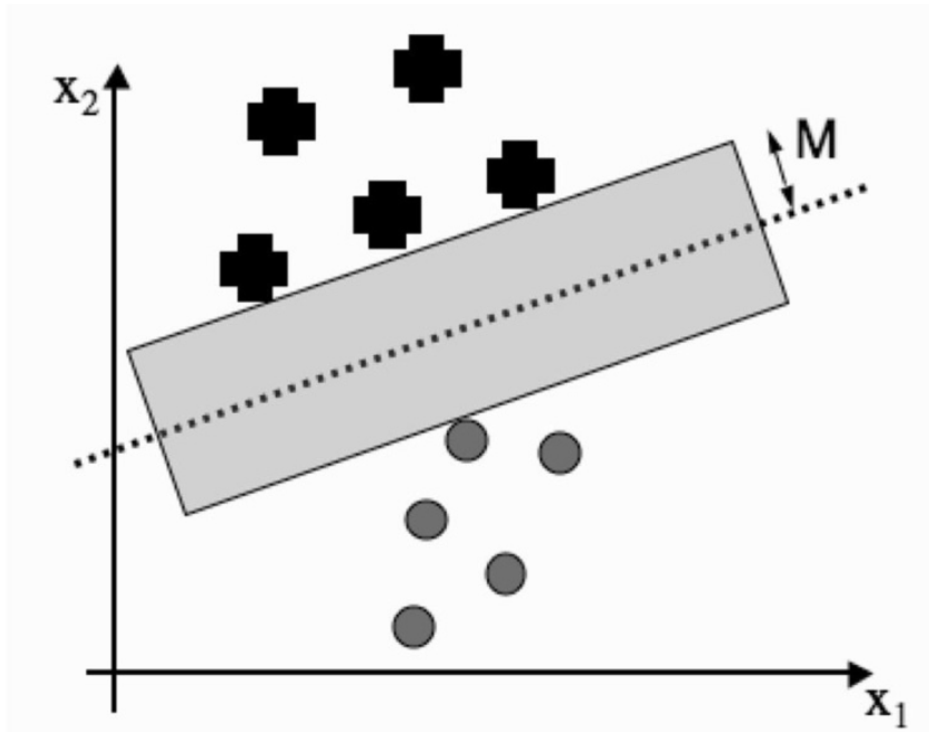
$t(\boldsymbol{w}'^T \boldsymbol{x} + w'_0) \geq 1$ Which is the canonical form

# Constraints

We consider a vector w valid only if:

$$\boxed{t\left(\boldsymbol{w}^{T}\boldsymbol{x}+w_{0}\right)\geq 1}$$

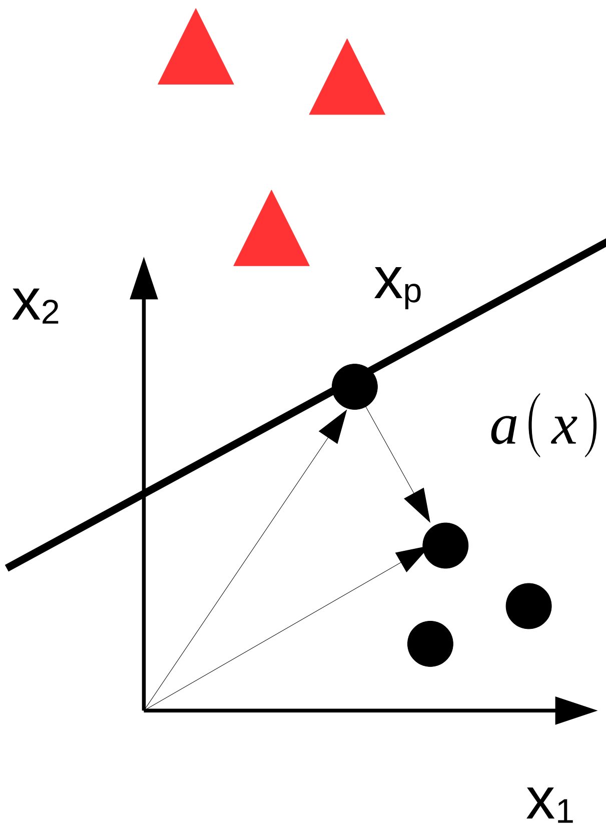
# The Margin




The margin is the distance between the closest point to the separation boundary, and the boundary itself.

# Recall from the perceptron…






Distance to the hyperplane

$$\boldsymbol{x} = \boldsymbol{x}_p + d\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}$$

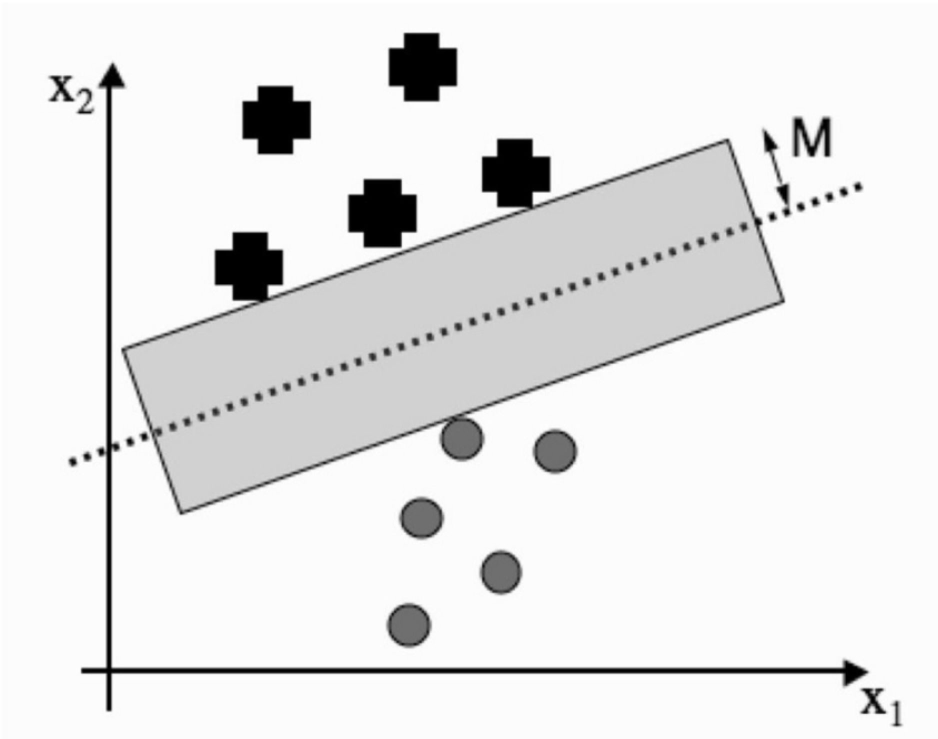
$$a(x) = \boldsymbol{w}\left(\boldsymbol{x}_p + d\frac{\boldsymbol{w}}{\|\boldsymbol{w}\|}\right) + w_0 = \underbrace{\boldsymbol{w}\,\boldsymbol{x}_p + w_0}_{=0} + d\frac{\boldsymbol{w}^T\boldsymbol{w}}{\|\boldsymbol{w}\|} = d\|\boldsymbol{w}\|$$

Because $x_p$ is on the hyperplane

$$\boldsymbol{w}^T\boldsymbol{x} + w_0 = d\|\boldsymbol{w}\|$$

# The Margin





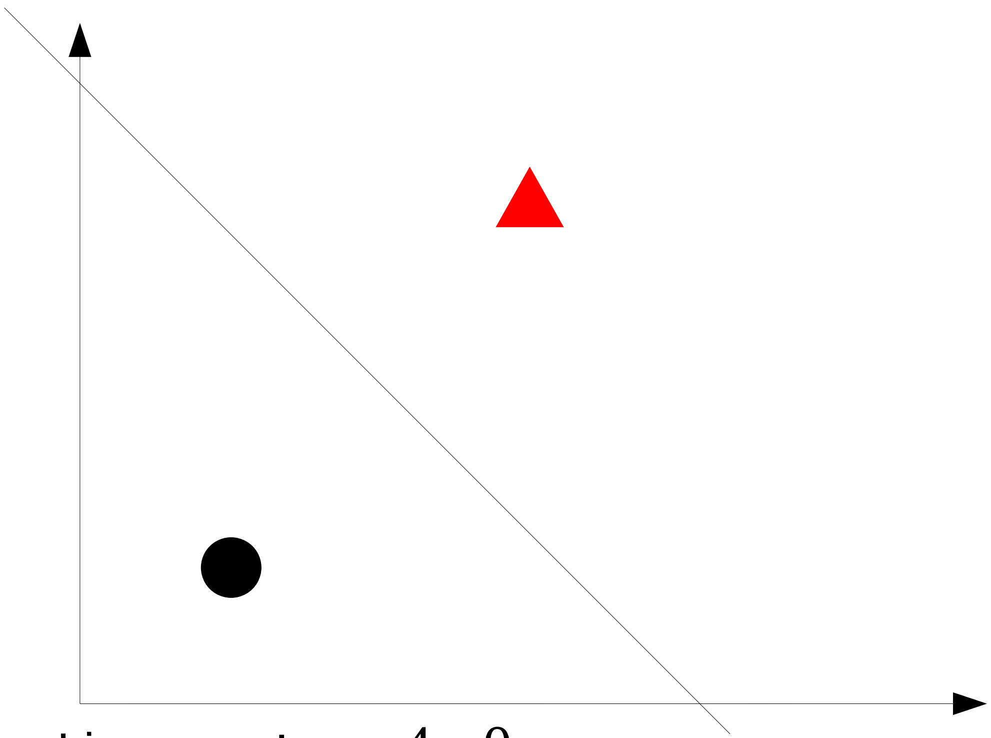

The margin is the distance between the closest point to the separation boundary, and the boundary itself

$$t = \{-1, +1\}$$

$$\frac{t\left(\boldsymbol{w}^T \boldsymbol{x} + w_0\right)}{\|\boldsymbol{w}\|} = |d|$$

# Let's look at this again...





Classes: {-1,1}

$$y(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{w}^T \boldsymbol{x} + w_0 > 0 \\ -1 & \text{if } \boldsymbol{w}^T \boldsymbol{x} + w_0 \leq 0 \end{cases}$$

$d\,||\boldsymbol{w}||$

Line: $x_1 + x_2 - 4 = 0$

point: $\langle 1, 1 \rangle$ desired class: $t = -1$ output: $1 + 1 - 4 = -2 \leq 0 \Rightarrow y = -1$

point: $\langle 3, 3 \rangle$ desired class: $t = 1$ output: $3 + 3 - 4 = 2 > 0 \Rightarrow y = 1$

# Maximum margin



We saw that for the closest points: $t(\boldsymbol{w}^T \boldsymbol{x} + w_0) = 1$

Therefore:

$$|d| = \frac{t(\boldsymbol{w}^T \boldsymbol{x} + w_0)}{\|\boldsymbol{w}\|} = \frac{1}{\|\boldsymbol{w}\|}$$

We can maximise it by minimising: $\|\boldsymbol{w}\|$

# The SVM Formulation

UNIVERSITY OF LEEDS

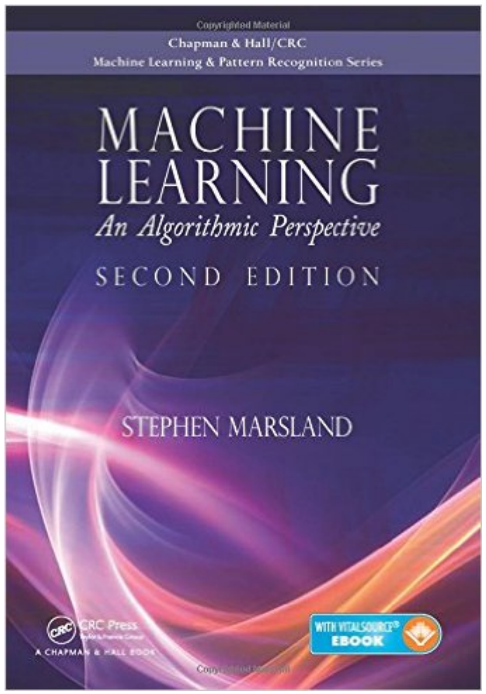
Margin as large
as possible

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

# Conclusion


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 8.1