Centre de Visió per Computador

Universitat Autònoma de Barcelona

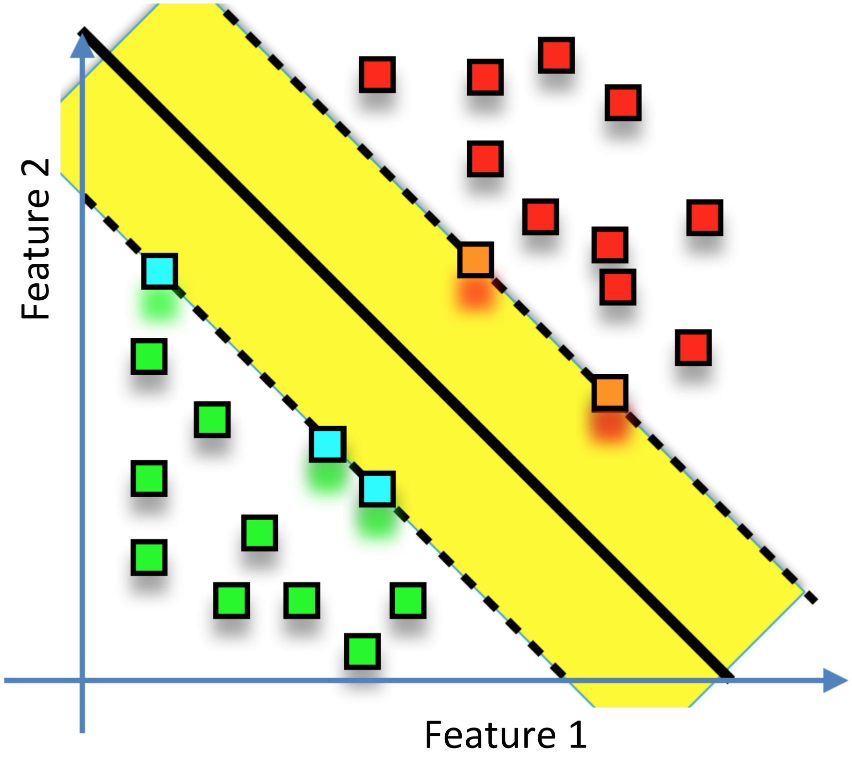
Master in Computer Vision
Barcelona

Module: M3. Machine Learning for Computer Vision

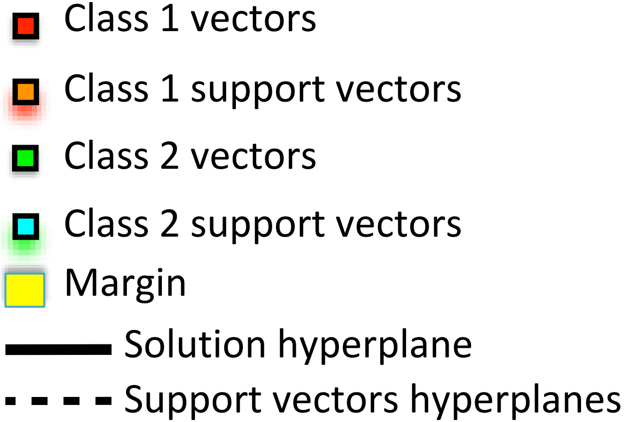
Lecture: The Support Vector Machine (SVM): Mathematical Development

Lecturer: Ramon Baldrich / Fernando Vilariño

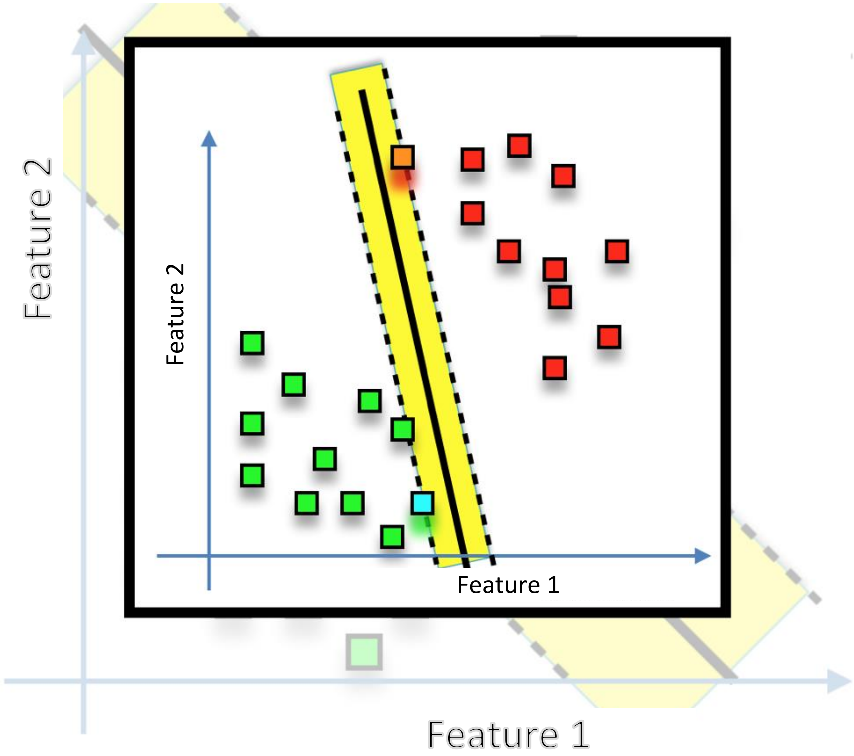
SVM: ***Linear** classifier based on **maximal margin** from **the support vectors.***



The separating hyperplane is obtained from the solution of an optimization problem: maximal distance between the 2 hyperplanes containing the support vectors of both classes (**maximal margin**).

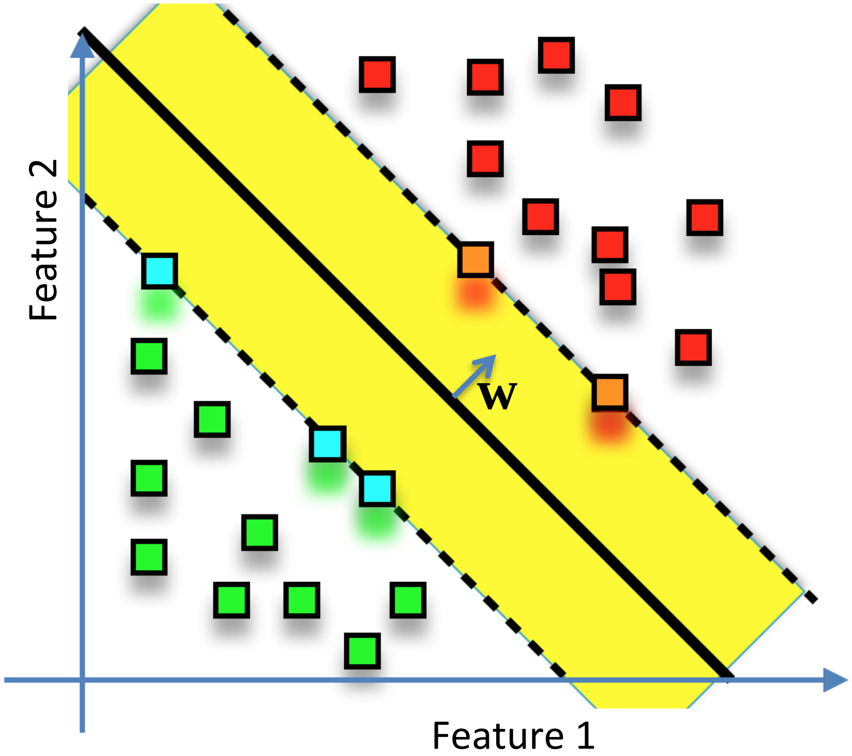


SVM:***Linear** classifier based on **maximal margin** from **the support vectors.***



Any other choice of support vectors generate a lower margin.

SVM: ***Linear*** *classifier based on* ***maximal margin*** *from* ***the support vectors.***



The samples of each class have one label:

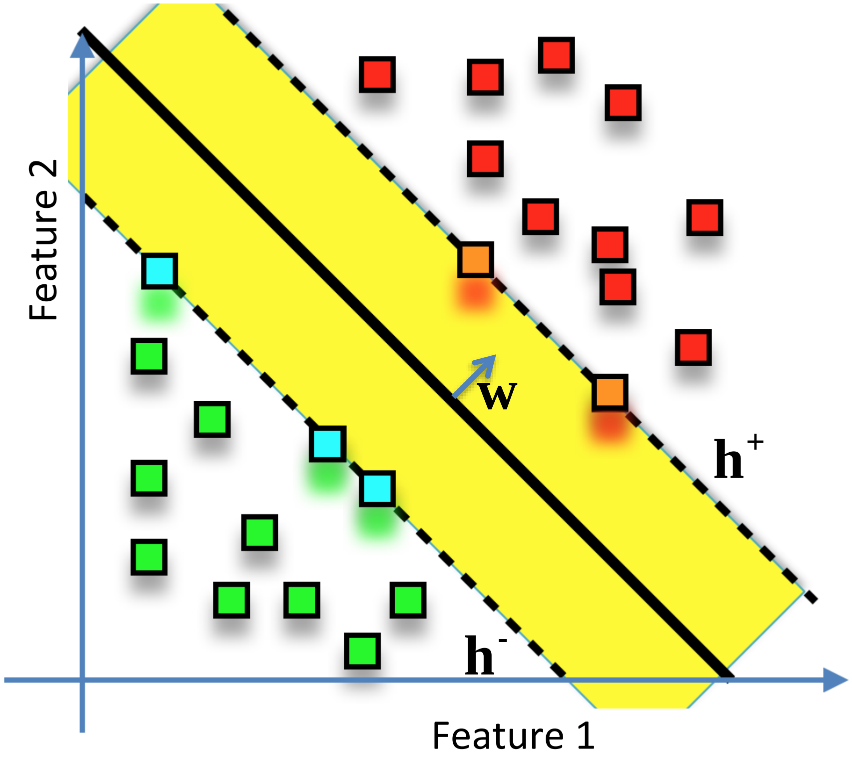
$y_1 = +1$

$y_2 = -1$

$$\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b = 0$$

Hyperplane solution $\mathbf{w}$

SVM: ***Linear*** *classifier based on* ***maximal margin*** *from* ***the support vectors.***



The samples of each class have one label:

$y_1 = +1$

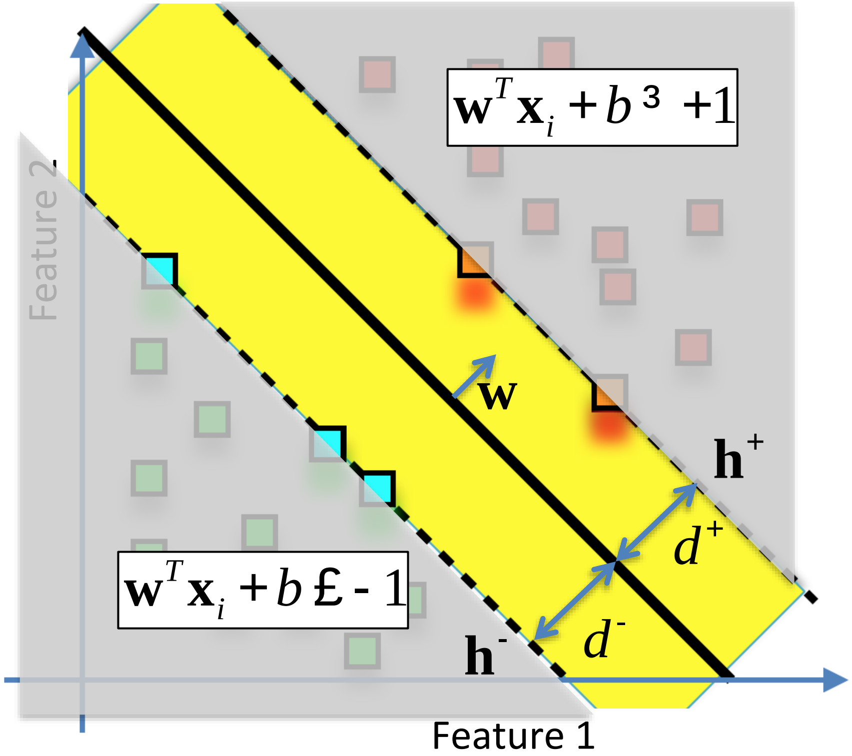
$y_2 = -1$

$$\mathbf{w}^{\mathrm{T}}\mathbf{x}_i + b = 0$$

$$\boxed{\begin{aligned} \mathbf{h}^{+} &\rightarrow \mathbf{w}^{T}\mathbf{x}_i + b = +1 \\ \mathbf{h}^{-} &\rightarrow \mathbf{w}^{T}\mathbf{x}_i + b = -1 \end{aligned}}$$

Support vector hyperplanes

SVM: ***Linear*** *classifier based on* ***maximal margin*** *from* ***the support vectors.***



The samples of each class have one label:

- $y_1 = +1$
- $y_2 = -1$

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

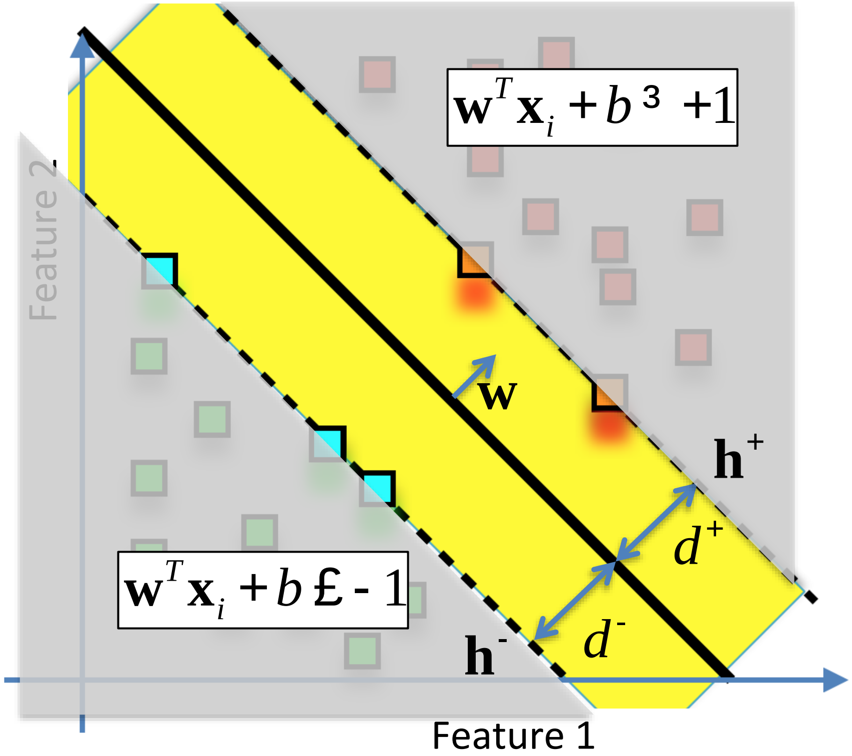
$$\mathbf{h}^+ \rightarrow \mathbf{w}^T\mathbf{x}_i + b = +1$$

$$\mathbf{h}^- \rightarrow \mathbf{w}^T\mathbf{x}_i + b = -1$$

Classification condition

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

SVM: ***Linear*** *classifier based on* ***maximal margin*** *from* ***the support vectors.***



The samples of each class have one label:

$y_1 = +1$

$y_2 = -1$

$$\mathbf{w}^\mathrm{T}\mathbf{x}_i + b = 0$$

$$\mathbf{h}^+ \rightarrow \mathbf{w}^T\mathbf{x}_i + b = +1$$

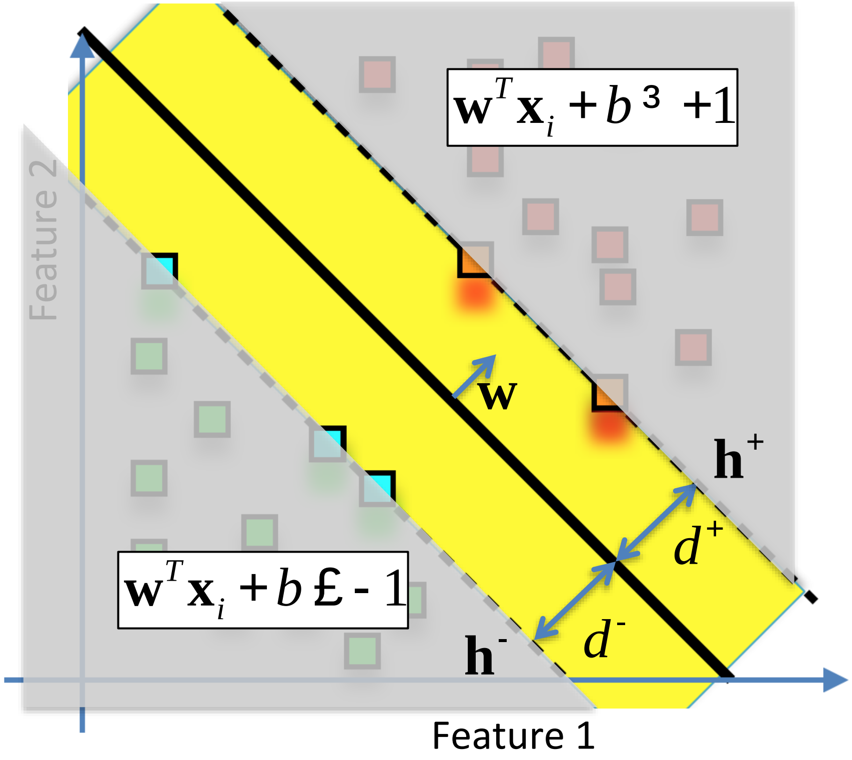
$$\mathbf{h}^- \rightarrow \mathbf{w}^T\mathbf{x}_i + b = -1$$

Classification condition

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

$$\boxed{d^+ = d^- = \frac{|\mathbf{wx}+b|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|} \qquad \text{margin} = d^+ + d^- = 2\frac{1}{\|\mathbf{w}\|}}$$

SVM: ***Linear*** *classifier based on* ***maximal margin*** *from* ***the support vectors.***



The samples of each class have one label:

$y_1 = +1$

$y_2 = -1$

$$\|\mathbf{w}\|^2 = \mathbf{w}^T\mathbf{w}$$

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

$$\mathbf{h}^+ \rightarrow \mathbf{w}^T\mathbf{x}_i + b = +1$$

$$\mathbf{h}^- \rightarrow \mathbf{w}^T\mathbf{x}_i + b = -1$$

$$\underset{\mathbf{w},b}{\mathrm{Minimise}}\, F(\mathbf{w}) = \frac{1}{2}\|\mathbf{w}\|^2$$

Subject to: $y_i(\mathbf{w}^T\mathbf{x}_i + b)$ ³ 1

SVM a quadratic optimisation problem

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\underset{\mathbf{w},b}{\text{Minimise}}\, \mathrm{F}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

Suject to: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\underset{\mathbf{w},b}{\text{Minimise}}\, \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

Suject to: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

Construction of the Lagrangian L:

$$L(x,\alpha) = f(x) + \sum_i \alpha_i g_i(x) \qquad \forall\ \alpha_i \geq 0$$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\underset{\mathbf{w},b}{\text{Minimise}}\, \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}$$

Suject to: $y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1$

Construction of the Lagrangian L:

$$L(x, \boxed{\alpha}) = \boxed{f(x)} + \sum_i \alpha \boxed{g_i(x)} \qquad \forall\ \alpha_i \geq 0$$

Function to optimise

Constraints

Lagrange multipliers

$$f(x) \rightarrow \Phi(\mathbf{w})$$
$$g_i(x) \rightarrow y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0$$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l} \underset{\mathbf{w},b}{\text{Minimise}}\, \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^{T}\mathbf{w} \\ \text{Suject to: } y_i(\mathbf{w}^{T}\mathbf{x}_i + b) \geq 1 \end{array}\right\} \Rightarrow \quad \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w}, b, \alpha)))$$

Construction of the Lagrangian L:

$$L(x, \boxed{\alpha}) = \boxed{f(x)} + \sum_i \alpha \boxed{g_i(x)} \qquad \forall\ \alpha_i \geq 0$$

Function to optimise

Constraints

Lagrange multipliers

$$f(x) \rightarrow \Phi(\mathbf{w})$$
$$g_i(x) \rightarrow y_i(\mathbf{w}^{T}\mathbf{x}_i + b) - 1 \geq 0$$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l} \underset{\mathbf{w},b}{\text{Minimise}}\, \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{Suject to: } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \end{array}\right\} \Rightarrow \quad \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w}, b, \alpha))$$

Construction of the Lagrangian L:

$$L(x, \boxed{\alpha}) = \boxed{f(x)} + \sum_i \alpha \boxed{g_i(x)} \qquad \forall\, \alpha_i \geq 0$$

Function to optimise

Constraints

Lagrange multipliers

$$f(x) \rightarrow \Phi(\mathbf{w})$$

$$g_i(x) \rightarrow y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0$$

The Lagrangian L of the SVM problem is:

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1]$$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l} \underset{\mathbf{w},b}{\text{Minimise}}\, \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{Suject to: } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \end{array}\right\} \Rightarrow \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w}, b, \alpha)))$$

Construction of the Lagrangian L:

$$L(x,\underbrace{\boxed{\alpha}}_{\text{Lagrange multipliers}}) = \underbrace{\boxed{f(x)}}_{\text{Function to optimise}} + \sum_i \alpha_i \underbrace{\boxed{g_i(x)}}_{\text{Constraints}} \qquad \forall\, \alpha_i \geq 0$$

$$f(x) \to \Phi(\mathbf{w})$$
$$g_i(x) \to y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0$$

The Lagrangian L of the SVM problem is:

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1]$$

The minimisation of L with respect to $\mathbf{w}, b$ implies:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \Rightarrow \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \Rightarrow \sum_i \alpha_i y_i = 0 \qquad \alpha_i \geq 0$$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l} \text{Minimise}_{\mathbf{w},b}\ \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{Suject to: } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \end{array}\right\} \Rightarrow \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w}, b, \alpha))$$

Construction of the Lagrangian L:

$$L(x, \boxed{\alpha}) = \boxed{f(x)} + \sum_i \alpha \boxed{g_i(x)} \qquad \forall\ \alpha_i \geq 0$$

Function to optimise

Constraints

Lagrange multipliers

$$f(x) \rightarrow \Phi(\mathbf{w})$$
$$g_i(x) \rightarrow y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0$$

The Lagrangian L of the SVM problem is:

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1]$$

The minimisation of L with respect to $\mathbf{w}, b$ implies:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$$

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\left(\sum_i \alpha_i y_i \mathbf{x}_i\right)^T \left(\sum_j \alpha_j y_j \mathbf{x}_j\right) - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j \mathbf{x}_j\right)^T \mathbf{x}_i + \sum_i \alpha_i =$$

$$= -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i$$

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l} \underset{\mathbf{w},b}{\text{Minimise}}\, \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{Suject to: } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \end{array}\right\} \Rightarrow \quad \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w}, b, \alpha))$$

Construction of the Lagrangian L:

$$L(x, \boxed{\alpha}) = \boxed{f(x)} + \sum_i \alpha \boxed{g_i(x)} \qquad \forall\, \alpha_i \geq 0$$

Function to optimise

Constraints

Lagrange multipliers

$$f(x) \rightarrow \Phi(\mathbf{w})$$
$$g_i(x) \rightarrow y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1 \geq 0$$

The Lagrangian L of the SVM problem is:

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1]$$

The minimisation of L with respect to $\mathbf{w}, b$ implies:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \qquad \Rightarrow \quad \sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$$

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\left(\sum_i \alpha_i y_i \mathbf{x}_i\right)^T \left(\sum_j \alpha_j y_j \mathbf{x}_j\right) - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j \mathbf{x}_j\right)^T \mathbf{x}_i + \sum_i \alpha_i =$$

$$= -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_j y_i \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i$$

$$\underset{\alpha}{\text{Maximize}}\ Q(\alpha) = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_j y_i \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i$$

Subject to: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

SVM DUAL

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l}\underset{\mathbf{w},b}{\text{Minimise}}\, F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w}\\ \text{Suject to: } y_i(\mathbf{w}^T\mathbf{x}_i+b)\geq 1\end{array}\right\} \Rightarrow \quad \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w},b,\alpha)))$$

$$\underset{\alpha}{\text{Maximise}}\, Q(\alpha) = -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i \alpha_i$$

Sujeto a: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

SVM DUAL

The Lagrangian L of the SVM problem is:

$$L(\mathbf{w},b,\alpha_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i+b)-1]$$

The minimisation of L with respect to $\mathbf{w}, b$ implies:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$$

$$L(\mathbf{w},b,\alpha_i) = \frac{1}{2}\left(\sum_i \alpha_i y_i \mathbf{x}_i\right)^T\left(\sum_j \alpha_j y_j \mathbf{x}_j\right) - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j \mathbf{x}_j\right)^T \mathbf{x}_i + \sum_i \alpha_i =$$

$$= -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i \alpha_i$$

$$\left.\begin{array}{l}\mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i\\ b = y_i - \mathbf{w}^T\mathbf{x}_i\end{array}\right\} \Rightarrow \quad f(\mathbf{x}) = \mathrm{sgn}\left(\sum_i y_i\alpha_i \mathbf{x}^T\mathbf{x}_i + b\right)$$

SVM Classifier

SVM: Mathematical development.

OBJECTIVE: Obtain $\mathbf{w}, b$ belonging to the solution hyperplane:

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

This is solved as a problem of quadratic optimisation:

$$\left.\begin{array}{l} \underset{\mathbf{w},b}{\text{Minimise}}\ \Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} \\ \text{Suject to: } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \end{array}\right\} \Rightarrow \max_{\alpha}(\min_{\mathbf{w},b}(L(\mathbf{w}, b, \alpha)))$$

SVM PRIMAL

$$\underset{\alpha}{\text{Maximise}}\ Q(\alpha) = -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i \alpha_i$$

Sujeto a: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

SVM DUAL

The Lagrangian L of the SVM problem is:

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\mathbf{w}^T\mathbf{w} - \sum_i \alpha_i[y_i(\mathbf{w}^T\mathbf{x}_i + b) - 1]$$

The minimisation of L with respect to $\mathbf{w}, b$ implies:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_i \alpha_i y_i \mathbf{x}_i = 0 \quad \Rightarrow \quad \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i$$

$$\frac{\partial L}{\partial b} = -\sum_i \alpha_i y_i = 0 \quad \Rightarrow \quad \sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$$

$$L(\mathbf{w}, b, \alpha_i) = \frac{1}{2}\left(\sum_i \alpha_i y_i \mathbf{x}_i\right)^T\left(\sum_j \alpha_j y_j \mathbf{x}_j\right) - \sum_i \alpha_i y_i\left(\sum_j \alpha_j y_j \mathbf{x}_j\right)^T\mathbf{x}_i + \sum_i \alpha_i =$$

$$= -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i \alpha_i$$

$$\left.\begin{array}{l} \mathbf{w} = \sum_i \alpha_i y_i \mathbf{x}_i \\ b = y_i - \mathbf{w}^T\mathbf{x}_i \end{array}\right\} \Rightarrow \quad f(\mathbf{x}) = \operatorname{sgn}\left(\sum_i y_i\alpha_i \mathbf{x}^T\mathbf{x}_i + b\right)$$

SVM Classifier

For the case of non linearly separable datasets: The **kernel trick**

$$\text{Maximise}_{\alpha}\ Q(\alpha) = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i$$

Subject to: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

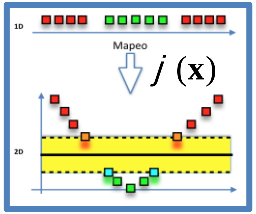
SVM DUAL

For the case of non linearly separable datasets: The **kernel trick**

$$\underset{\alpha}{\text{Maximise}}\ Q(\alpha) = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i$$

Subject to: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

SVM DUAL



Mapping function:

$\mathbf{x} \mapsto \varphi(\mathbf{x})$

Kernel function:

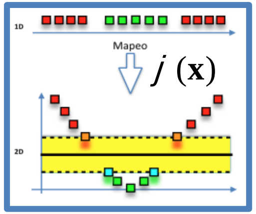
$K(\mathbf{x},\mathbf{z}) = \varphi(\mathbf{x})^T \varphi(\mathbf{z})$

For the case of non linearly separable datasets: The **kernel trick**

$$\text{Maximise}_{\alpha}\ Q(\alpha) = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_i \alpha_i$$

Subject to: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

SVM DUAL



Mapping function:

$\mathbf{x} \mapsto \varphi(\mathbf{x})$

Kernel function:

$K(\mathbf{x}, \mathbf{z}) = \varphi(\mathbf{x})^T \varphi(\mathbf{z})$

$$\text{Maximise}_{\alpha}\ Q(\alpha) = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \mathbf{x}_j) + \sum_i \alpha_i$$

$$b = y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) = y_i - \sum_j y_j \alpha_j K(\mathbf{x_j}, \mathbf{x_i})$$

$$f(\mathbf{x}) = \text{sgn}\left(\sum_i y_i \alpha_i K(\mathbf{x}, \mathbf{x_i}) + b\right)$$

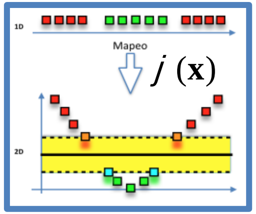
Kernel SVM

For the case of non linearly separable datasets: The **kernel trick**

$$\underset{\alpha}{\text{Maximise}}\ Q(\alpha) = -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j + \sum_i \alpha_i$$

Subject to: $\sum_i \alpha_i y_i = 0 \quad \alpha_i \geq 0$

SVM DUAL



Mapping function:

$\mathbf{x} \mapsto \varphi(\mathbf{x})$

Kernel function:

$K(\mathbf{x},\mathbf{z}) = \varphi(\mathbf{x})^T\varphi(\mathbf{z})$

$$\underset{\alpha}{\text{Maximise}}\ Q(\alpha) = -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j K(\mathbf{x}_i\mathbf{x}_j) + \sum_i \alpha_i$$

$$b = y_i - \mathbf{w}^T\varphi(\mathbf{x}_i) = y_i - \sum_j y_j\alpha_i K(\mathbf{x_j},\mathbf{x_i})$$

$$f(\mathbf{x}) = \text{sgn}\Big(\sum_i y_i\alpha_i K(\mathbf{x},\mathbf{x_i}) + b\Big)$$

Kernel SVM

Examples of kernel functions:

Polinomial: $K(\mathbf{x},\mathbf{z}) = \langle \mathbf{x},\mathbf{z}\rangle^d$

Radial basis functions: $K(\mathbf{x},\mathbf{z}) = e^{-\|x-z\|^2/2\sigma}$

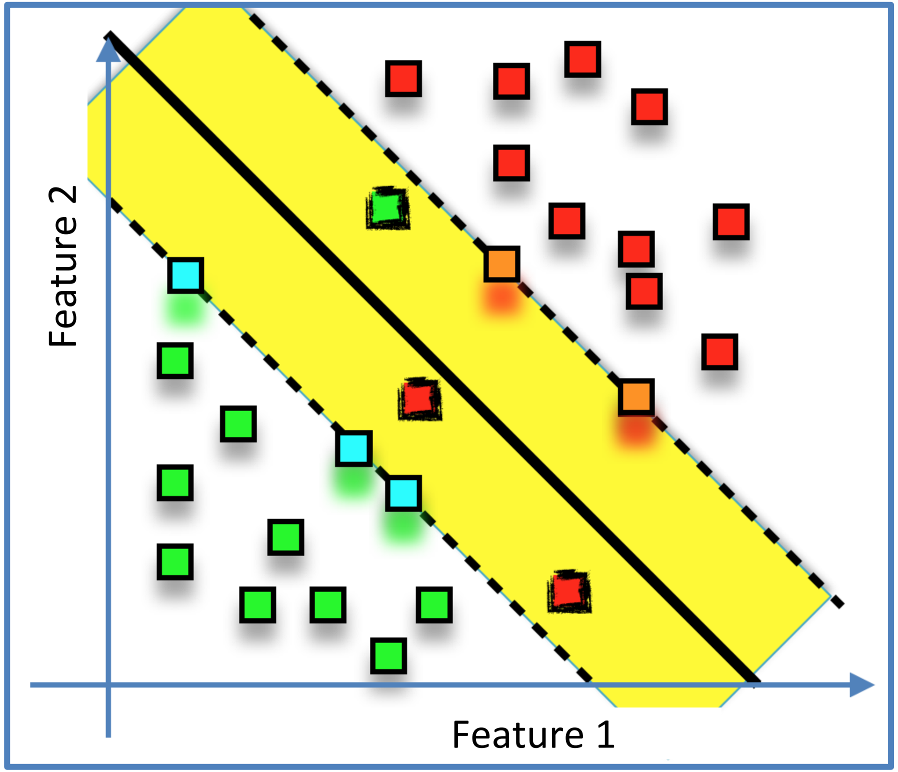
Sigmoid: $K(\mathbf{x},\mathbf{z}) = \tanh(\kappa\langle \mathbf{x},\mathbf{z}\rangle - \delta)$

Multi-inverse quadratic: $K(\mathbf{x},\mathbf{z}) = (\|\mathbf{x}-\mathbf{z}\|^{1/2}2\sigma + c^2)^{-1}$

Intersection kernel: $K(\mathbf{x},\mathbf{z}) = \sum_{i=1}^{n}\min(x(i),z(i))$

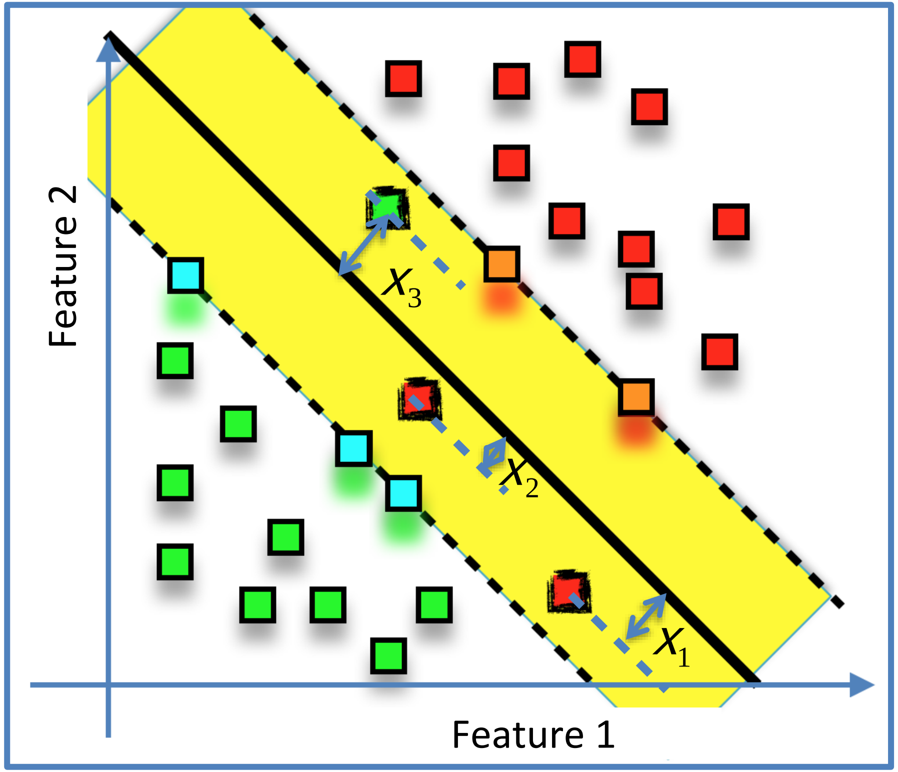
Kernel functions introduce new parameters the value of which must be fixed, generally, through cross-validation.

Non linearly separable datasets: **Soft margin**.



Vectors violating the margin condition

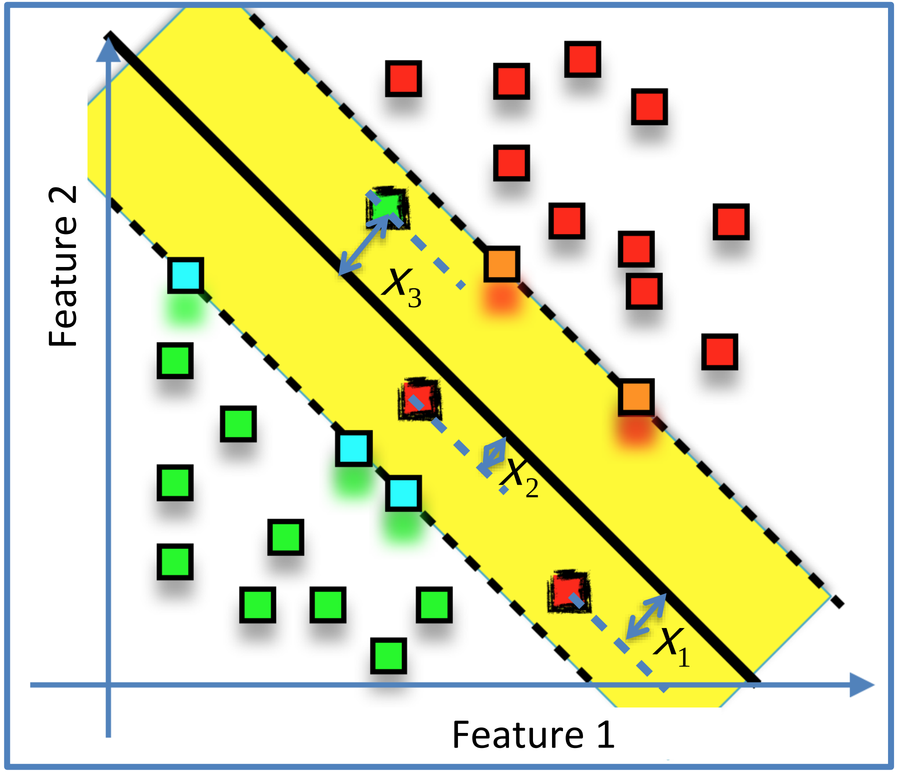
Non linearly separable datasets: **Soft margin**.



Vectors violating the margin condition

Slack variables: $\xi_i \geq 0$

Non linearly separable datasets: **Soft margin.**



Vectors violating the margin condition

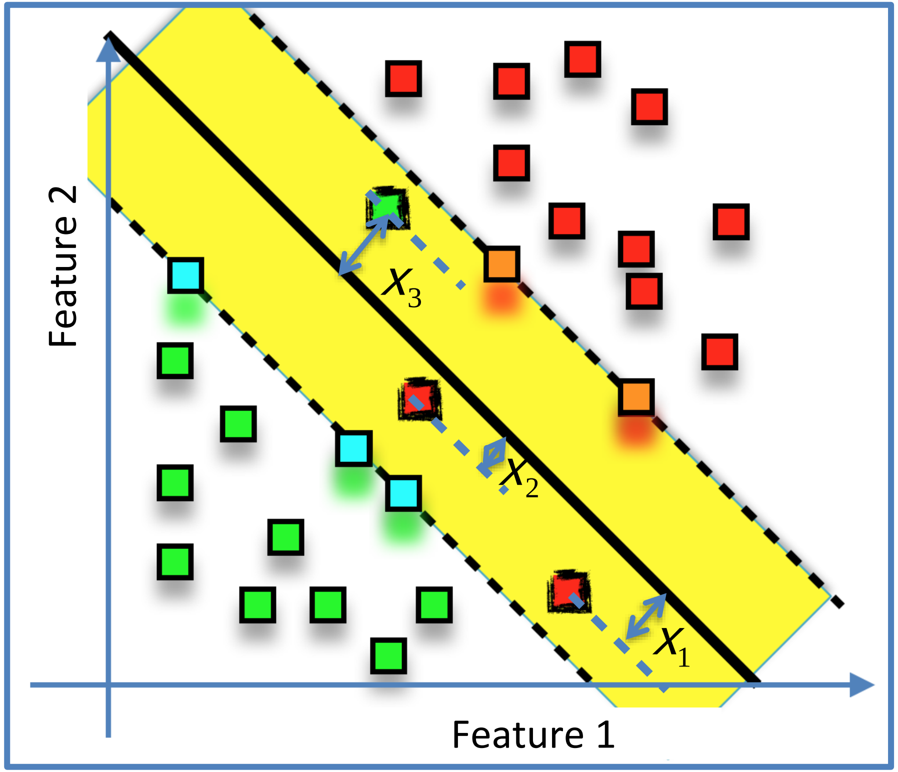
Slack variables: $\xi_i \geq 0$

$$\underset{\mathbf{w},b,\xi}{\text{Minimise}}\; F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i$$

Subject to: $y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i$

SVM PRIMAL

Non linearly separable datasets: **Soft margin.**



Vectors violating the margin condition

Slack variables: $\xi_i \geq 0$

$$\underset{\mathbf{w},b,\xi}{\text{Minimise}}\ F(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i$$

Subject to: $y_i(\mathbf{w}^T f(\mathbf{x}_i) + b) \geq 1 - \xi_i$

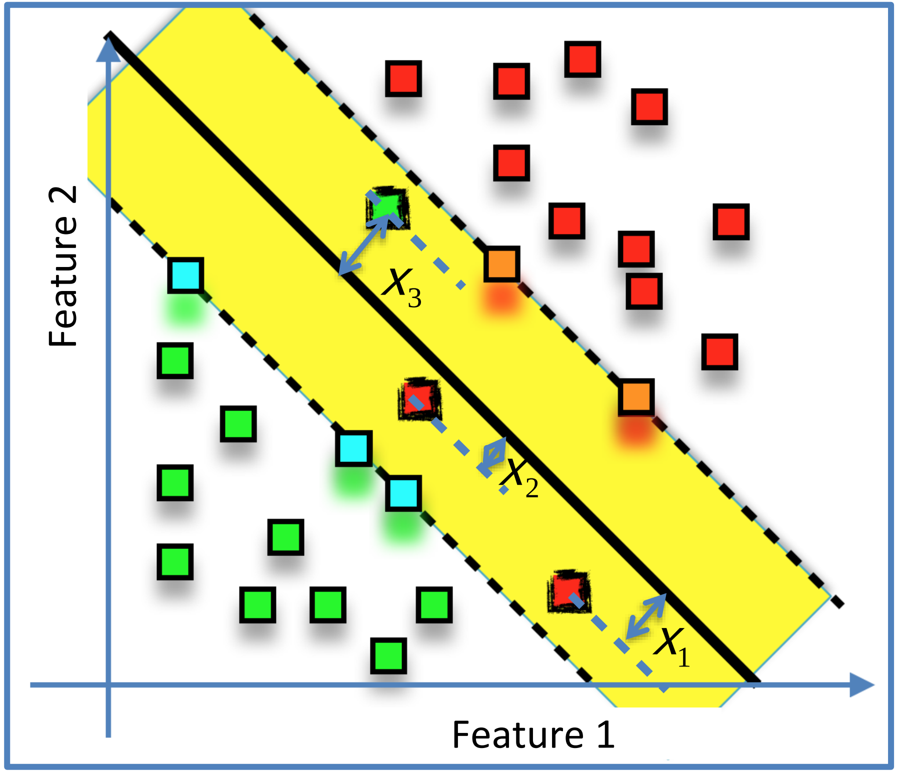
SVM PRIMAL

$$\underset{\alpha}{\text{Maximise}}\ Q(\alpha) = -\frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i \mathbf{x}_j) + \sum_i \alpha_i$$

Subject to: $\begin{cases} 0 \leq \alpha_i \leq C \\ \sum_i \alpha_i y_i = 0 \end{cases}$

SVM DUAL

Non linearly separable datasets: **Soft margin.**



Vectors violating the margin condition

Slack variables: $\xi_i \geq 0$

$$\underset{\mathbf{w},b,\xi}{\text{Minimise}}\ \mathrm{F}(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_i \xi_i$$

Subject to: $y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i$

SVM PRIMAL

$$\underset{\alpha}{\text{Maximise}}\ \mathrm{Q}(\alpha) = -\frac{1}{2}\sum_i\sum_j \alpha_i\alpha_j y_i y_j K(\mathbf{x}_i\mathbf{x}_j) + \sum_i \alpha_i$$

Subject to:
$$\begin{cases} 0 \leq \alpha_i \leq C & b = y_i(1-\xi_i) - \sum_j y_j \alpha_i K(\mathbf{x_j}, \mathbf{x_i}) \\ \sum_i \alpha_i y_i = 0 & f(\mathbf{x}) = \operatorname{sgn}(\sum_i y_i \alpha_i K(\mathbf{x}, \mathbf{x_i}) + b) \end{cases}$$

SVM DUAL

Key concepts:

- SVM as a ***quadratic optimisation problem***.
- The SVM solution as a ***dual*** problem.
- The dual problem allows the explicit solution from ***scalar products***.
- The solution in form of scalar products allow the introduction of ***kernels***.
- Kernel functions allow to classify ***non linearly separable datasets***.
- The **slack variables** introduce a regularisation factor that allows tolerance to errors by relaxing the condition of maximal margin.
- Both the kernel parameters and the regularisation factor m**ust be adjusted during the training stage**.