Support vector machines (SVM's) - Vapnik

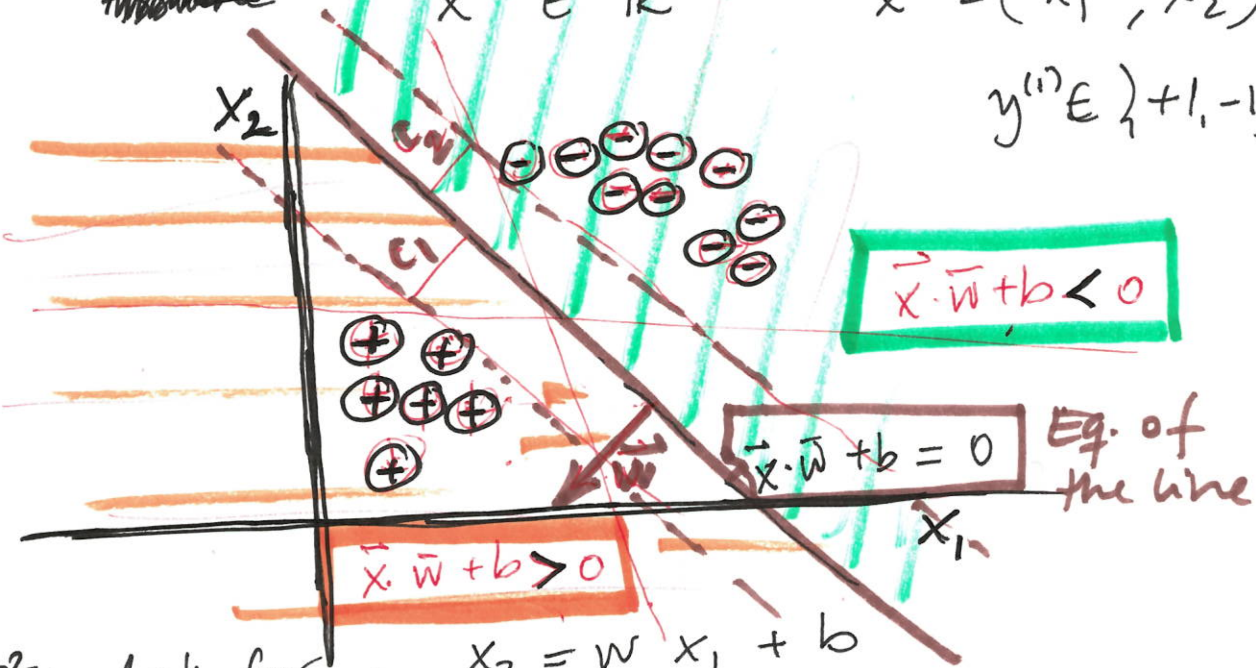
Classification as geometric separation.

$$\left[\ \vec{x}^{(i)} = (x_1^{(i)}, \dots, x_p^{(i)}) \text{ with } y^{(i)} \in \{-1, +1\}\ \right]$$

<u>Consider 2D case first:</u> ~~Assume~~

$\vec{x}^{(i)} \in \mathbb{R}^2$ $\qquad \vec{x}^{(i)} = (x_1^{(i)}, x_2^{(i)})$

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



In $\mathbb{R}^2$: look for a line : $x_2 = \underline{w}\, x_1 + \underline{b}$

$$\vec{x} \cdot \vec{w} + b = 0$$

$$\vec{w} = (w, -1)$$

$$(w, -1) \cdot (x_1, x_2) + b = 0 \ \checkmark$$

In general,

for $\vec{x} \in \mathbb{R}^p$

Note that:

$$\left[\vec{x}^T \cdot \vec{\beta} \quad = \right] \vec{x} \cdot \vec{w} + b = 0$$

Equation of a hyperplane

$$\begin{pmatrix} 1 \\ \vec{x} \end{pmatrix} = \vec{x} = \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_p \end{pmatrix} \quad \vec{\beta} = \begin{pmatrix} b \\ w_1 \\ \vdots \\ w_p \end{pmatrix} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_p \end{pmatrix} \quad \vec{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix}$$

Find the hyperplane

$$\left| \begin{array}{l} \vec{x} \cdot \vec{w} + b > 0 \text{ below} \\ \vec{x} \cdot \vec{w} + b < 0 \text{ above} \end{array} \right\} \qquad \vec{x} \cdot \underline{\underline{\vec{w}}} + \underline{\underline{b}} = 0$$

$(\vec{w}, b)$ to be found.

$$c_1, c_2 \in \mathbb{R}^+ \quad \left\{ \begin{array}{ll} \vec{x}^{(i)} \cdot \vec{w} + b > c_1 & \text{if } y^{(i)} = +1 \\ \vec{x}^{(i)} \cdot \vec{w} + b < -c_2 & \text{if } y^{(i)} = -1 \end{array} \right\}$$

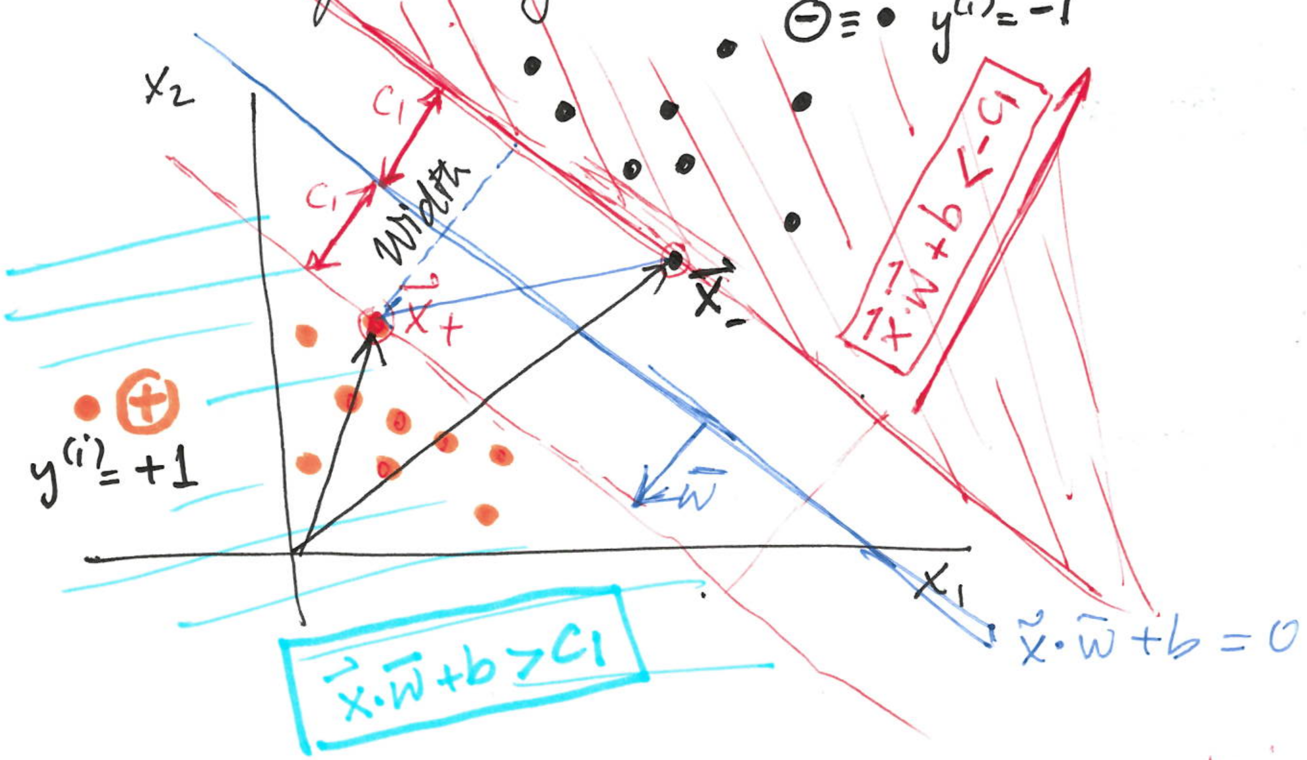
Once we have $(\vec{w}, b) \equiv \vec{\beta}^T$, we have our model to classify according to the above

$$\text{Given } \vec{x}^{in} \quad \left\{ \begin{array}{ll} \vec{x}^{in} \cdot \vec{w} + b > 0 & \text{then } \hat{y} = +1 \\ \vec{x}^{in} \cdot \vec{w} + b < 0 & \text{then } \hat{y} = -1 \end{array} \right.$$

What is the criterion for the hyperplane?

Make the width of the street as large as possible

Geometry of the SVM criterion:



① $$y^{(i)}\left(\vec{x}^{(i)}\cdot\vec{w}+b\right) \geq 0$$

Condition on both sides of hyperplane

② The solution is half way:

$$\vec{x}_+\cdot\vec{w}+b=c_1$$

$$\vec{x}_-\cdot\vec{w}+b=-c_1$$

$$\text{width}=\left(\vec{x}_+-\vec{x}_-\right)\cdot\frac{\vec{w}}{\|\vec{w}\|}=\frac{2c_1}{\|\vec{w}\|}$$

$\boxed{c_1=1}$ Can always be fixed

$$\boxed{\text{width}=\frac{2}{\|\vec{w}\|}}$$

max width $\Rightarrow$ $\min \frac{1}{2}\|\vec{w}\|^2$

$\boxed{\begin{array}{l} \text{Optimization (SVM):} \\ \min_{\vec{w}} \ \frac{1}{2}\|\vec{w}\|^2 \\ \text{subject to} \quad \underbrace{y^{(i)}(\vec{x}^{(i)} \cdot \vec{w} + b) \geq 1}_{} \quad i = 1, \dots, N \end{array}}$

All points above/below the two lines at distance 1 from $\vec{x} \cdot \vec{w} + b = 0$

Reminder :

(1) Lagrange multipliers :

Optimise a function subject to equality constraints

$$\boxed{\begin{array}{l} \min_{\vec{x}} \ f(\vec{x}) \\ \text{subject to} \quad g_i(\vec{x}) = 0 \quad i = 1, \dots, m \end{array}}$$

Lagrangian : $L = f(x) + \sum_{i=1}^{m} \alpha_i \, g_i(\vec{x})$

$$\left.\begin{cases} \nabla_{\vec{x}} L = 0 \\ \nabla_{\alpha} L = 0 \end{cases}\right\} \text{at } (\vec{x}^*, \vec{\alpha}^*)$$

(2) Linear programming:

The constraints define a convex set

Optimise a linear objective function under linear constraints (equalities and inequalities)

$$\min_{\vec{x}} \ \vec{w} \cdot \vec{x} \quad \text{(linear cost)}$$

subject to $g_i(\vec{x}) = 0 \quad i = 1, \ldots, m$

g_i are linear

$h_j(\vec{x}) \geq 0 \quad j = 1, \ldots, k$

h_j are linear

Simplex solves this problem.

(3) Quadratic program: (QP)

Solution given by KKT

(Karush 1939, Kuhn 1951, Tucker)

$$\min_{\vec{x}} \ f(\vec{x}) \quad \text{quadratic}$$

subject to linear $g_i(x) \leq 0$

$i = 1, \ldots, m$

Construct Lagrangian :   Lagrange multipliers

$$L = f(\vec{x}) + \sum_{i=1}^{m} \alpha_i \, \underbrace{g_i(\vec{x})}$$

At the minimum : $\vec{z}^*$, there exists $\vec{\alpha} = (\alpha_1, \dots, \alpha_m)$ such that :

Primal

① $\quad -\nabla f(\vec{z}^*) = \sum_{i=1}^{m} \alpha_i \nabla g_i(\vec{z}^*)$, $\quad \nabla L\big|_{z^*} = 0$

② $\quad g_i(\vec{z}^*) \leq 0 \quad \forall i \qquad \nabla L\big|_{\vec{\alpha}} = 0$

Dual

③ $\quad \alpha_i \geq 0 \quad \forall i$

④ $\quad \alpha_i \cdot g_i(\vec{z}^*) = 0 \quad \forall i$

Complementary slackness.

Applied to SVM optimisation :

$$\boxed{\begin{array}{l} \min\limits_{\bar{w}} \ \frac{1}{2} \|\bar{w}\|^2 \\ \text{subject to } \ \underbrace{1 - y^{(i)}(\vec{x}^{(i)} \cdot \bar{w} + b)}_{g_i(\vec{w}, b) \quad i = 1, \dots, N} \leq 0 \end{array}}$$

Construct Lagrangian:

$$\mathbf{L}(\vec{w}, b; \vec{\alpha}) = \frac{1}{2}\|\vec{w}\|^2 + \sum_{i=1}^{N} \alpha_i \left(1 - y^{(i)}(\vec{x}^{(i)} \cdot \vec{w} + b)\right)$$

$$\& \quad \frac{\partial L}{\partial b} = - \sum_{i=1}^{N} y^{(i)} \alpha_i$$

$$\nabla_{\vec{w}} L = \vec{w} - \sum_{i=1}^{N} \alpha_i y^{(i)} \vec{x}^{(i)}$$

At the minimum they equal zero

$$\left\{ \begin{array}{l} \sum_{i=1}^{N} y^{(i)} \alpha_i = 0 \\ \vec{w}^* = \sum_{i=1}^{N} \alpha_i y^{(i)} \vec{x}^{(i)} \end{array} \right\}$$

Rewrite in terms of $\vec{\alpha}$:

$$L(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i y^{(i)} \vec{x}^{(i)} \cdot \vec{w} -$$

$$- \cancel{\sum_{i=1}^{N} \alpha_i y^{(i)} b} + \frac{1}{2} \vec{w} \cdot \vec{w}$$

$$L(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \vec{w} \cdot \vec{w} = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y^{(i)} y^{(j)} \vec{x}^{(i)} \cdot \vec{x}^{(j)}$$

Conclusion: $\left[\alpha_i = 0 \quad \text{if} \quad \vec{x}^{(i)} \neq \vec{x}_+ \text{ or } \vec{x}_- \right]$

From ④ $\quad \alpha_i \left(1 - y^{(i)} (\vec{x}^{(i)} \cdot \vec{w} + b)\right) = 0 \quad \forall i$

$$\vec{w}^* = \sum_{i=1}^{N} \underline{\underline{\alpha_i}}\, y^{(i)} \vec{x}^{(i)}$$

$$= \underline{\underline{\alpha_+}}\, \vec{x}_+ + \underline{\underline{\alpha_-}}\, \vec{x}_-$$

"Anchor points"

Support vectors

These are found by the process of convex optimization above

Hard-margin SVM

(Perfect separation)

How do we extend SVMs to more realistic scenarios when there is no hyperplane that can separate the classes perfectly?

① Soft-margin SVM ≡ soft boundary

② Going beyond linear. ⇒ Kernels