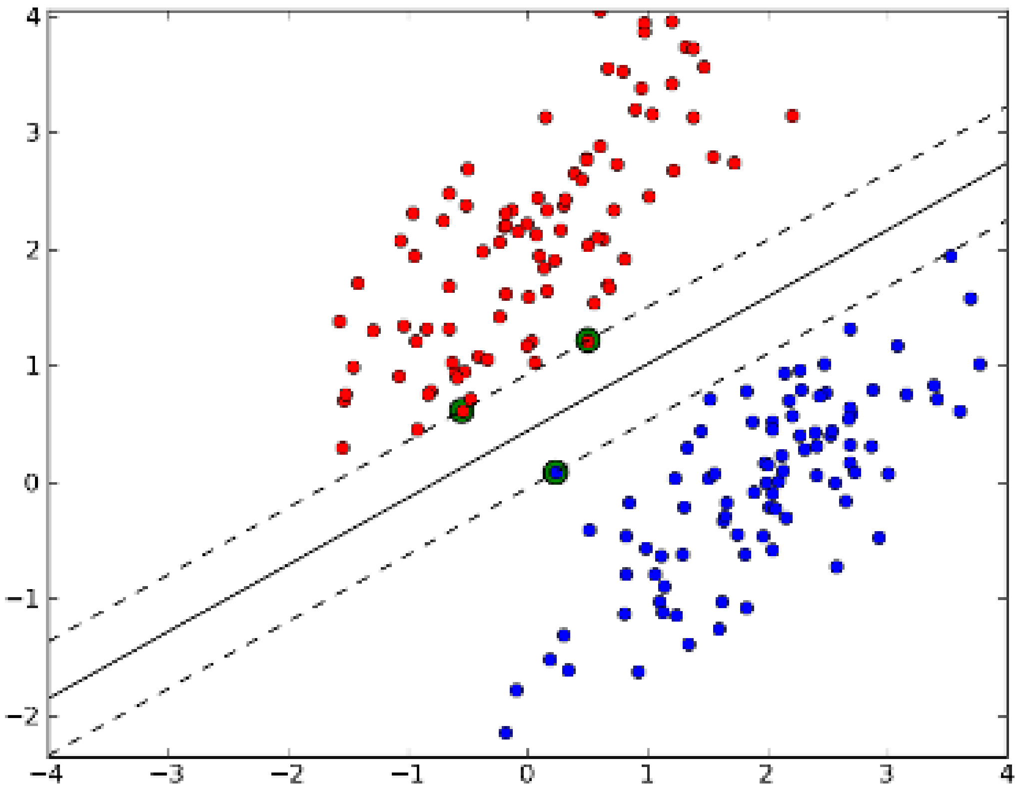
When the data is **linearly separable**, Hard SVM finds the linear separator with maximum margin.



Suppose $\{(\boldsymbol{x_1}, y_1), \ldots, (\boldsymbol{x_m}, y_m)\}$ is a set of labeled vectors that are **linearly separable**.

Hard SVM
$\text{minimize} \frac{\lVert \boldsymbol{w} \rVert^2}{2}$ such that $y_i\left(\boldsymbol{x}_i^T \boldsymbol{w}\right) \geq 1$

Suppose $\{(\boldsymbol{x_1}, y_1), \ldots, (\boldsymbol{x_m}, y_m)\}$ is a set of labeled vectors that are **linearly separable**.

Hard SVM is equivalent to:

$$\max_{\boldsymbol{\alpha} \in \mathbb{R}^m, \boldsymbol{\alpha} \geq \mathbf{0}} \left\{ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \left(\boldsymbol{x}_i^T \boldsymbol{x_j} \right) \right\}$$

$$\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x_i}$$

Definition: Kernel

The function $K: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a kernel if it can be written as an inner product:

- There exists a mapping $\Phi: \mathcal{X} \rightarrow \mathbb{R}^d$ such that $K(x, y) = \Phi(x)^T \Phi(y)$ for all $x, y \in \mathcal{X}$.

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^m,\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\left(\boldsymbol{x}_i^T\boldsymbol{x}_{\boldsymbol{j}}\right)\right\}$$

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^m,\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\left(\Phi(\boldsymbol{x_i})^T\Phi(\boldsymbol{x_j})\right)\right\}$$

$$\max_{\boldsymbol{\alpha}\in\mathbb{R}^m,\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\left(\Phi(\boldsymbol{x_i})^T\Phi(\boldsymbol{x_j})\right)\right\}$$

$$=\max_{\boldsymbol{\alpha}\in\mathbb{R}^m,\boldsymbol{\alpha}\geq\mathbf{0}}\left\{\sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j K\left(\boldsymbol{x_i},\boldsymbol{x_j}\right)\right\}$$

To classify a new example $\boldsymbol{x}$:

$$\boldsymbol{w}^T \Phi(\boldsymbol{x})$$

To classify a new example $\boldsymbol{x}$:

$$\boldsymbol{w}^T \Phi(\boldsymbol{x})$$

$$= \left(\sum_{i=1}^{m} \alpha_i y_i \Phi(\boldsymbol{x_i}) \right)^T \Phi(\boldsymbol{x})$$

To classify a new example $\boldsymbol{x}$:

$$\boldsymbol{w}^T \Phi(\boldsymbol{x})$$

$$= \left(\sum_{i=1}^{m} \alpha_i y_i \Phi(\boldsymbol{x_i})\right)^T \Phi(\boldsymbol{x})$$

$$= \sum_{i=1}^{m} \alpha_i y_i \, \Phi(\boldsymbol{x_i})^T \Phi(\boldsymbol{x})$$

To classify a new example $\boldsymbol{x}$:

$$\boldsymbol{w}^T \Phi(\boldsymbol{x})$$

$$= \left(\sum_{i=1}^{m} \alpha_i y_i \Phi(\boldsymbol{x_i})\right)^T \Phi(\boldsymbol{x})$$

$$= \sum_{i=1}^{m} \alpha_i y_i \, \Phi(\boldsymbol{x_i})^T \Phi(\boldsymbol{x})$$

$$= \sum_{i=1}^{m} \alpha_i y_i K(\boldsymbol{x_i}, \boldsymbol{x})$$

Question.

Let A be a **positive semidefinite matrix**.

($A = UU^T$ for some matrix U.)

Why is $K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T A \boldsymbol{x}'$ is a legal kernel?

Question.

Let A be a **positive semidefinite matrix**.

($A = U^T U$ for some matrix U.)

Why is $K(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{x}^T A \boldsymbol{x}'$ is a legal kernel?

Answer.

Let $\Phi(\boldsymbol{x}) = U\boldsymbol{x}$. Then $\Phi(\boldsymbol{x})^T \Phi(\boldsymbol{x}') = \boldsymbol{x}^T U^T U \boldsymbol{x}' = \boldsymbol{x}^T A \boldsymbol{x}' = K(\boldsymbol{x}, \boldsymbol{x}')$.

Theorem

A symmetric function $K: \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a kernel if and only if it is positive semidefinite.

In other words, K is a kernel if and only if for all $x_1, \dots, x_m \in \mathcal{X}$, the matrix $G_{i,j} = K(x_i, x_j)$ is a positive semidefinite matrix.