When the data is **linearly separable**, Hard SVM finds the linear separator with maximum margin.
Suppose \( \{(x_1, y_1), \ldots, (x_m, y_m)\} \) is a set of labeled vectors that are **linearly separable**.

**Hard SVM**

\[
\text{minimize} \quad \frac{\|w\|^2}{2} \\
\text{such that} \quad y_i(x_i^T w) \geq 1
\]
Suppose \{ (x_1, y_1), \ldots, (x_m, y_m) \} is a set of labeled vectors that are **linearly separable**.

Hard SVM is equivalent to:

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

\[
w = \sum_{i=1}^{m} \alpha_i y_i x_i
\]
**Definition: Kernel**

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

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

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

\[
= \max_{\alpha \in \mathbb{R}^m, \alpha \geq 0} \left\{ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right\}
\]
To classify a new example $x$:

$$w^T \Phi(x)$$
To classify a new example $x$: 

$$w^T \Phi(x) = \left( \sum_{i=1}^{m} \alpha_i y_i \Phi(x_i) \right)^T \Phi(x)$$
To classify a new example $x$:

$$w^T \Phi(x)$$

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

$$= \sum_{i=1}^{m} \alpha_i y_i \Phi(x_i)^T \Phi(x)$$
To classify a new example $x$:  
\[ w^T \Phi(x) \]

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

\[
= \sum_{i=1}^{m} \alpha_i y_i \Phi(x_i)^T \Phi(x) 
\]

\[
= \sum_{i=1}^{m} \alpha_i y_i K(x_i, x) 
\]
Question.

Let $A$ be a positive semidefinite matrix.

\[(A = UU^T \text{ for some matrix } U.)\]

Why is $K(x, x') = x^T Ax'$ is a legal kernel?
Question.

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

\[ A = U^T U \text{ for some matrix } U. \]

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

Answer.

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

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 $x_1, \ldots, x_m \in \mathcal{X}$, the matrix $G_{i,j} = K(x_i, x_j)$ is a positive semidefinite matrix.