we will not discuss more mathematical details as it will go into the loop of derivation with the Calculus; but I will give you the end result.

It turns out the classifier problem is solved by the inner product of two vectors (two observation).

Inner product of 2 'r' vectors 'a' and 'b' is defined as

$\langle a,b \rangle = \sum_{i=1}^{r} a_i b_i$.

$$\langle x_i, x_{i'} \rangle = \sum_{j=1}^{p} x_{ij} x_{i'j}.$$

using this a linear support vector classifier can be represented as

$$f(x) = \beta_0 + \sum_{i=1}^{n} \alpha_i \langle x, x_i \rangle$$

when there are n parameters α_i, $i = 1, \ldots n$, one per training observation.

To estimate the parameters $\alpha_1, \ldots \alpha_n$ and β_0 and all we need are the $\binom{n}{2}$ inner products $\langle x_i, x_i' \rangle$ between all pairs of observation $\frac{n(n-1)}{2}$.

Now to do a generalization we multiply a kernel function $K(x_i, x_i')$.

where K is the kernel function.

$$\boxed{K(x_i, x_i') = \sum_{j=1}^{p} x_{ij} x_{i'j}.}$$ → linear Kernel.

$$\boxed{f(x) = \beta_0 + \sum_{i \in S} \alpha_i K(x, x_i)}$$ → Polynomial Kernel.

$\boxed{\text{Most used Kernel - Radial basis function (RBF) function.}}$