A version of the support vector classifier optimization problem is for some positive $C^*$

$$\min_{\beta \in \mathbb{R}^p, \beta_0 \in \mathbb{R}} \frac{1}{2} \|\beta\|^2 + C^* \sum_{i=1}^{N} \xi_i \text{ subject to } \begin{cases} y_i (x_i' \beta + \beta_0) + \xi_i \geq 1 & \forall i \\ \text{for some } \xi_i \geq 0 \end{cases}$$

where ultimately the SV classifier is

$$\hat{f}(x) = \text{sign}[x' \beta + \beta_0]$$

Or equivalently the optimization problem is for some positive $C$

$$\max_{\mathbf{u} \in \mathbb{R}^p \text{ with } \|\mathbf{u}\| = 1, \beta_0 \in \mathbb{R}} M \text{ subject to } \begin{cases} y_i (x_i' \mathbf{u} + \beta_0) \geq M (1 - \xi_i) & \forall i \\ \text{for some } \xi_i \geq 0 \text{ with } \sum_{i=1}^{N} \xi_i \leq C \end{cases}$$

where ultimately the SV classifier is

$$\hat{f}(x) = \text{sign}[x' \mathbf{u} + \beta_0]$$

For a given $C^*$ there is a corresponding $C$ and the optimizing $\mathbf{u}$ in the second formulation is the unit vector corresponding to the optimizing $\beta$ in the first formulation, $\beta/\|\beta\|$, while the $\beta_0$ in the second formulation is $\beta_0$ in the first divided by $\|\beta\|$ from the first. The margin (defining the "cushion around classification boundary" is $M$ in the second formulation and $1/\|\beta\|$ in the first. Points with optimized $\xi_i \geq 0$ are "support vectors," ones with optimized $\xi_i = 0$ being "on the margin" around the classification boundary, ones with $\xi_i > 0$ being on the "wrong side" of the appropriate margin and ones with $\xi_i > 1$ indicating that the training set point $x_i$ is misclassified by the SV classifier. Both $C^*$ and $C$ are complexity parameters, small $C^*$ and large $C$ corresponding to low complexity classifiers.

The notion of a non-negative definite kernel $K(x, z)$ and an implicit abstract linear space to which $\mathbb{R}^p$ is mapped via some (abstract) function $T(x)$ and in which the inner product of $T(x)$ and $T(z)$ is $K(x, z)$ provides a very flexible generalization of support vector classifiers.

One might declare that what is sought is some element of the abstract space of the form

$$S = \sum_{i=1}^{N} \beta_i T(x_i)$$

(i.e. some constants $\beta_1, \beta_2, \ldots, \beta_N$) and some constant $\beta_0$ that minimize for some positive $C^*$

$$\frac{1}{2} \|S\|^2 + C^* \sum_{i=1}^{N} \xi_i$$

subject to

$$y_i ([T(x_i), S] + \beta_0) + \xi_i \geq 1 \text{ for some } \xi_i \geq 0$$

where ultimately the classifier will be of the form

$$\hat{f}(x) = \text{sign}([T(x), S] + \beta_0)$$

(The inner product here is the inner product in the abstract space corresponding to $K(x, z)$ and the norm is the norm in that space produced by the inner product.) This amounts to saying that we’ll think of solving the support vector classifier problem in the abstract space and then effectively doing classification by moving the data to that space via the (abstract) transformation $T(x)$. 

1
The elements of this are easily expressed in terms of the kernel. First

\[ \|S\|^2 = \langle S, S \rangle \]

\[ = \left( \sum_{i=1}^{N} \beta_i T(x_i), \sum_{i=1}^{N} \beta_i T(x_i) \right) \]

\[ = \sum_{i,j} \beta_i \beta_j \langle T(x_i), T(x_j) \rangle \]

\[ = \sum_{i,j} \beta_i \beta_j K(x_i, x_j) \]

\[ = \beta' K \beta \]

for \( \beta' = (\beta_1, \beta_2, \ldots, \beta_N) \in \mathbb{R}^N \) and \( K \) the \( N \times N \) matrix with \( i,j \) entry \( K(x_i, x_j). \) Then

\[ \langle T(x), S \rangle = \sum_i \beta_i K(x, x_i) \]

\[ = (K(x, x_1), K(x, x_2), \ldots, K(x, x_N)) \beta \]

\[ = \kappa(x)^T \beta \]

for an \( N \)-vector-valued function \( \kappa \) with \( l \)th entry \( \kappa(x)_l = K(x, x_l). \)

So the optimization problem is for some positive \( C^* \) to minimize (the quadratic in \( \beta \) plus penalty on the slack variables \( \xi_i \))

\[ \frac{1}{2} \beta' K \beta + C^* \sum_{i=1}^{N} \xi_i \]

subject to the constraint

\[ y_i (\kappa(x_i)^T \beta + \beta_0) + \xi_i \geq 1 \quad \forall i \text{ for some } \xi_i \geq 0 \]

which is essentially of the same form as the optimization problem corresponding to a support vector classifier where now the "data vectors" are \( \kappa(x_i) \) (rather than \( x_i \) as in the untransformed problem). Upon solving the optimization problem (that is finding a support vector classifier in the abstract space based on a solution of a dual problem in \( \mathbb{R}^N \)), the ultimate corresponding classifier in the original variable \( x \) is

\[ \hat{f}(x) = \text{sign} [\kappa(x)^T \beta + \beta_0] \]

\[ = \text{sign} \left[ \sum_i \beta_i K(x, x_i) + \beta_0 \right] \]

and one has built a "voting function" (a surrogate for the likelihood ratio or conditional probability of class 1 given \( x \) if the classifier is any good) using a linear combination of slices of the kernel function taken at input vectors \( x_i \) in the training data set.

In this formulation, the margin is \( M = 1/\sqrt{\beta' K \beta} \) in the abstract space around the set of points with inner product with \( S \) equal to \(-\beta_0 \). In terms of points in \( \mathbb{R}^p \), the two surfaces representing points \( M \) (in the abstract space) away from the classification boundary

\[ \left\{ x \in \mathbb{R}^p \mid \sum_i \beta_i K(x, x_i) + \beta_0 = 0 \right\} \]

are of the forms

\[ \left\{ x \in \mathbb{R}^p \mid \sum_i \beta_i K(x, x_i) + \beta_0 = 1 \right\} \text{ and } \left\{ x \in \mathbb{R}^p \mid \sum_i \beta_i K(x, x_i) + \beta_0 = -1 \right\} \]

Training vectors \( x_i \in \mathbb{R}^p \) correspond to support vectors in the abstract space if they are "on the wrong side" of their respective margins around the linear classification boundary in the abstract space. (Points \( x_i \) with \( y_i = -1 \) are support vectors if \( \sum_i \beta_i K(x_i, x_i) + \beta_0 > -1 \) and points \( x_i \) with \( y_i = 1 \) are support vectors if \( \sum_i \beta_i K(x_i, x_i) + \beta_0 < 1 \). These support vectors turn out to correspond to \( i \) with \( \beta_i \neq 0 \) )