I: Convex hulls and linear separability

Answer:

The Convex hull of S is given by,

$$S = \{x : x = \sum_{i=1}^{m} \alpha_i x_i \text{ , } where\ \alpha_i \geq 0\ \forall\ i \in\ [m], \sum_{i=1}^{m} \alpha_i = 1\}$$

$$S = (x_1 \ldots\ldots x_m) = x_i$$

Consider another set of D dimensional observation

$$T = (y_1 \ldots\ldots y_m) = y_i$$

So, the Convex hull will be

$$(T) = \left\{y : y = \sum_{i=1}^{m} \beta_i\, y_i \geq 0 \forall\ i \in [m], \sum_{i=1}^{m} \beta_i = 1\right\}$$

If the 2 points set should be linearly separable then there should exist a vector $\widehat{w}$ and a scalar w_0 such that

$$\widehat{w_T} S + w_0 > 0\ \forall\ S \text{ , and}$$

$$\widehat{w_T} T + w_0 < \ 0\ \forall\ T\ .$$

We must show that the convex hull of the above two points sets do not intersect. So, the linear discriminant for S:

$$z(x) = \widehat{w_T} S + w_0$$

Now substituting (1) In (3):

$$z(x) = \widehat{w_T}\left(\sum_{i=1}^{m} \alpha_i\ S\right) + w_0$$

Because α_i is a scalar quantity

$$= \sum_{i=1}^{m} \alpha_i\ (\widehat{w_T} S) + w_0$$

$$= \sum_{i=1}^{m} \alpha_i(\ \widehat{w_T} S + w_0) \qquad \ldots \text{(I)}$$

Similarly, for another set of D dimensional observation:

$$z(y) = \sum_{i=1}^{m} \beta_i\ \ (\widehat{w_T} T + w_0) \qquad \ldots \text{(II)}$$

Where, $\beta_i \geq 0$ and $\sum_{i=1}^{m} \beta_i = 1$

There must exist a point which is common if the both convex hulls intersects between x_i and y_i. The linear discriminant of that point will be from equation (I) and (II)

$$z(xy) = \ \sum_{i=1}^{m} \alpha_i\ (\widehat{w_T} S + w_0) = \sum_{i=1}^{m} \beta_i\ \ (\widehat{w_T} T + w_0) \qquad \ldots \text{(III)}$$

But for linear separability,

$$z(\ x_m) = \widehat{w_T}S + w_0 > 0,$$

$$z\left(\ y_m\right) = \widehat{w_T}T + w_0 < \ 0 \qquad \ldots \text{(IV)}$$

The above equations lead to contradiction. It is not possible in equation (III) that the equations are simultaneously greater and less than zero. We can state that if their Convex hull do not intersect then the two sets of points (S and T) are linearly separable.

III: Non-equivalence of "hard" and "soft" SVM

Answer:

Claim mentioned in question is wrong as the algorithm won't return the exact hypothesis.

Let us consider some integer $i\ >\ 1$ and j $>\ 0$. Now let $x_0 = (0, \alpha) \epsilon\, R^2$, where $a\, \epsilon\, (0,1)$.

For $a = 1, \cdots, i - 1$ and let $x_a = (0, a)$.

Let $y_0 \cdots y_{i-1} = 1$. $S = \left\{(x_m, y_m) : m\, \epsilon \{0,1, \cdots, i - 1\}\right\}$

The solution of hard SVM is $w = (0,1/\alpha)$.

If,

$$\text{j}.1\ + \frac{1}{i}(1 - \alpha) \leq \frac{1}{\alpha^2}$$

The solution of soft SVM is $k = (0,1)$. Since, $a\, \epsilon\, (0,1)$ such that $\frac{1}{\alpha^2} > \text{j} + \frac{1}{i}$.

We can observe that there exists $\alpha_0 > 0$ then for every $\alpha_0 >\ \alpha$ it will hold the desired inequality.

If α is small enough then soft -SVM neglects x_0 .

IV: Kernel construction

Answer: Given that $K_1$and K_2 are valid kernels.

A valid kernel function should satisfy two property i.e. Symmetry Property and Positive Semi Definiteness property. We restrict them to finite set of points. Therefore, K_1andK_2 become kernel matrices. Consider finite set of points $x_1, \ldots, x_m$

$\mathrm{K}(u, v) = \alpha K_1(u, v) + \beta K_2(u, v)$. K_1 is a Positive Semi-Definite. …(I)

As α is positive therefore the Positive Semi-Definiteness of αK_1 is always positive. The Gram Matrix of this kernel will remain same as α is multiplied to each element of the matrix.
αK_1 is a valid kernel because the properties of a valid kernel are satisfied.
Similarly we can prove that βK_2 is also a valid kernel.

From (I) Let $\mathrm{K} = \alpha \mathrm{K}_1 + \beta \mathrm{K}_2$
$\alpha K_1$and βK_2 are symmetric gram matrices as proved earlier. Symmetric gram matrix is obtained by adding two symmetric gram matrices. The Positive Semi-Definiteness of the both matrices are positive.
As it is a linear combination, the combined matrix also has Positive Semi-Definiteness which is positive. The properties of a valid kernel are satisfied. Therefore,
$K(u, v) = \alpha K_1(u, v) + \beta K_2(u, v)$ is a valid kernel.

$\mathrm{K}(u, v) = K_1(u, v) * K_2(u, v)$

From Mercer's theorem, $K_1$and K_2 are valid kernels which must have an inner product representation. Let a and b denote feature vectors of $K_1$and K_2 respectively. Then,

$$K_1(x, y) = a(x)^T a(y), ia(z) = [a_1(z), \ldots a_M(z)]$$
$$K_2(x, y) = b(x)^T b(y), b(z) = [b_1(z), \ldots b_N(z)]$$

a is a function that produces an M-dimensional vector, and b produces an N-dimensional vector.

$$K(x, y) = K_1(x, y) K_2(x, y)$$
$$= \left(\textstyle\sum_{m=1}^{M} a_m(x) a_m(y)\right)\left(\textstyle\sum_{n=1}^{N} b_n(x) b_n(y)\right)$$
$$= \textstyle\sum_{m=1}^{M} \sum_{n=1}^{N} [a_m(x) b_n(x)][a_m(y) b_n(y)]$$

Now let $f(z) = M.N - dimensional\ vector$, such that

$$c_{m_n}(z) = a_m(z) b_n(z) = \sum_{m=1}^{M} \sum_{n=1}^{N} f_{mn}(x) f_{mn}(y)$$
$$= f(x)^T f(y)$$

Therefore, by using Mercer's theorem we proved that $K(u, v)$ can be expressed in terms of inner product using feature map f. K is a valid kernel.