Support Vector Machines Problems

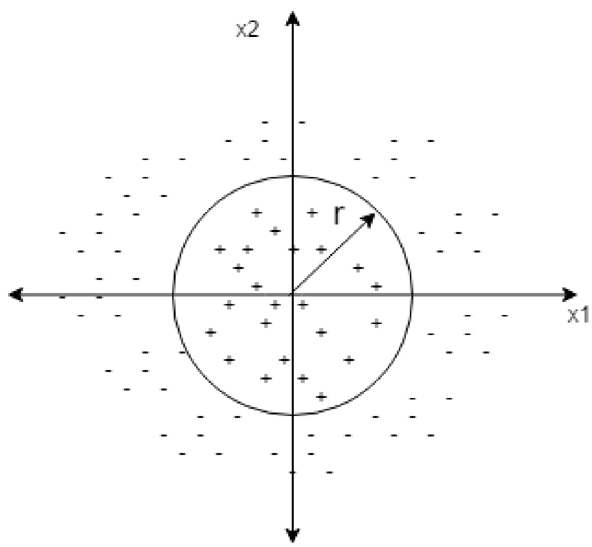
CS 536 Machine Learning
Nitin Reddy Karolla

March 13, 2019

1. Suppose you had a data set in two dimensions that satisfied the following: the positive class all lay within a certain radius of a point, the negative class all lay outside that radius.

   — Show that under the feature map $\phi(x_1, x_2) = (1, x_1, x_2, x_1 x_2, x_2^2, x_2^2)$ (or equivalently, with the kernel $K(x, y) = (1 + x.y)^2$), a linear separator can always be found in this embedded space, regardless of radius and where the data is centered.

Figure 1: Distribution of data



As seen in above plot, the positive class all lay within a certain radius (r) of a point, the negative class all lay outside that radius.

We now have the equation of the above circle as:

$$x_1^2 + x_2^2 = r^2$$

The above equation can be rearranged to get the following:

$$T = r^2 - x_1^2 + x_2^2$$

Now, we can define a function,

$$f(\underline{X}) = sign(T) = \begin{cases} \text{-1} & \text{if } T > 0 \\ +1 & \text{if } T < 0 \end{cases}$$

Using the polynomial kernel, we can now transform the current data space into feature space of higher dimension. Polynomial kernel obtained is there following:

$$\phi(x_1, x_2) = (1, x_1, x_2, x_1x_2, x_2^2, x_2^2)$$

As per the above transformation, there exist a linear separator in the feature space of the form $\phi((x_1, x_2) = (1, x_2^2, x_2^2)$. This linear separator can always be found and, is independent of the radius and where the data is centered.

— In fact show that if there is an ellipsoidal separator, regardless of center, width, orientation (and dimension!), a separator can be found in the quadratic feature space using this kernel.

The ellipse can be represented using the below equation -

$$a(x_1 - z_1)^2 + b(x_2 - z_2)^2 = 1$$

where a is the major axis, b is the minor axis,(z_1, z_2) is the center of the ellipse.

On equating above equation to T, we can rewrite it as :

$$T = 1 - ax_1^2 + 2az_1x_1 - az_1^2 - bx_2^2 + 2bz_2x_2 - bz_2^2$$
$$T = (1 - az_1^2 - bz_2^2) + 2az_1x_1 + 2bz_2x_2 - ax_1^2 - bx_2^2$$

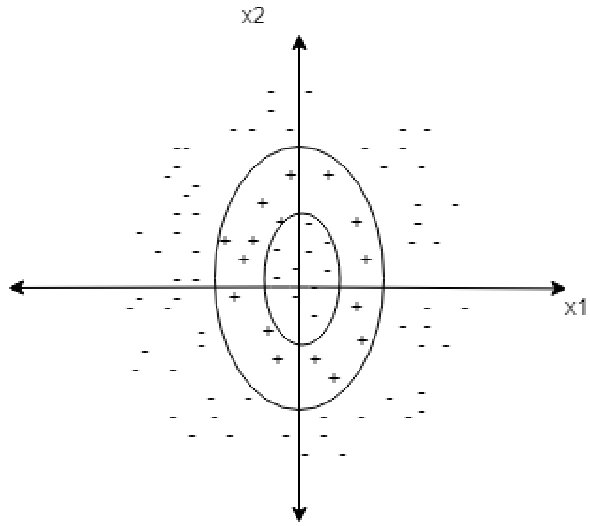
Like in previous problem, we can transform the current data space into the feature of higher dimension using polynomial kernel.

$$\phi(x_1, x_2) = (1, x_1, x_2, x_1x_2, x_2^2, x_2^2)$$

As per the above transformation, there exist a linear separator in the feature space of the form $\phi((x_1, x_2) = (1, x_1, x_2, x_2^2, x_2^2)$. As we can see, regardless of the center, width, orientation, a separator can be found in the quadratic feature space using the polynomial kernel.

2. As an extension of the previous problem, suppose that the two dimensional data set satisfied the following: the positive class lay within one of two (disjoint) ellipsoidal regions, and the negative class was everywhere else. Argue that the kernel $K(x, y) = (1 + x.y)^4$ will recover a separator.

Figure 2: Distribution of data



As seen in the above figure, the data is separated in such a way that negative class is within the smaller ellipse and outside the bigger ellipse, while the positive class is between the smaller and the bigger ellipse.

Let the center of both ellipse be at (z_1, z_2) and the major axes of both ellipses are a_1 and a_2, similary the minor axes are b_1 and b_2. The equations of two ellipses are the following:

$$T_1 = 1 - (\frac{(x_1 - z_1)^2}{a_1^2} + \frac{(x_2 - z_2)^2}{b_1^2})$$
$$T_2 = 1 - (\frac{(x_1 - z_1)^2}{a_2^2} + \frac{(x_2 - z_2)^2}{b_2^2})$$

Given the above two equations, the data can be split into classes based on the following criteria :

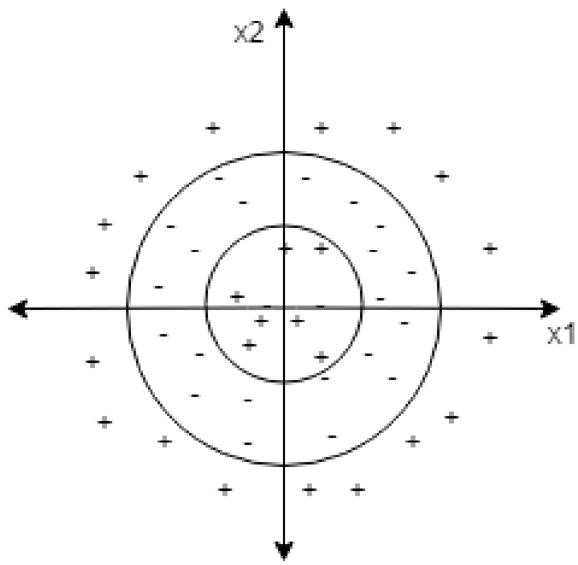
$$f(\underline{X}) = -sign(T_1.T_2) = \begin{cases} \text{-1}, & \text{if } T_1 > 0 \text{ and } T_2 > 0 \\ +1, & \text{if } T_1 < 0 \text{ and } T_2 > 0 \\ \text{-1}, & \text{if } T_1 < 0 \text{ and } T_2 < 0 \end{cases}$$

To obtain this, we would need the data space that will be take all features into consideration arsing out of the multiplication of T_1 and T_2. Clearly we can see that multiplication of T_1 and T_2 is in order of x_1^4, x_2^4.

Hence, a kernel of form $K(\underline{x}, \underline{y}) = (1 + \underline{xy})^4$ is required to find a separator for the above data.

3. Suppose that the two dimensional data set is distributed like the following: the positive class lays in a circle centered at some point, the negative class lies in a circular band surrounding it of some radius, and then additional positive points lie outside that radius. Argue that the kernel $K(x, y) = (1 + x.y)^4$ will recover a separator.

Figure 3: Distribution of data



As seen in the above figure, the data is separated in such a way that positive class is within the smaller circle and outside the bigger circle, while the negative class is between the smaller and the bigger circle.

Let the radius of both circle be at (r_1, r_2). The equations of two circles are the following:

$$T_1 = r_1^2 - x_1^2 - x_2^2$$
$$T_2 = r_2^2 - x_1^2 - x_2^2$$

Given the above two equations, the data can be split into classes based on the following criteria :

$$f(\underline{X}) = sign(T_1.T_2) = \begin{cases} +1, & \text{if } T_1 > 0 \text{ and } T_2 > 0 \\ \text{-1}, & \text{if } T_1 < 0 \text{ and } T_2 > 0 \\ +1, & \text{if } T_1 < 0 \text{ and } T_2 < 0 \end{cases}$$

On solving $T_1.T_2$, we have :

$$\begin{aligned} T_1.T_2 &= (r_1^2 - x_1^2 - x_2^2)(r_2^2 - x_1^2 - x_2^2) \\ &= (r_1 r_2)^2 - (r_1^2 + r_2^2)x_1^2 - (r_1^2 + r_2^2)x_2^2 + 2(x_1 x_2)^2 + x_1^4 + x_2^4 \end{aligned}$$

As we see above, the $T_1.T_2$ is in the order of x_1^4 and $x2^4$.

Hence, for the transformation of data, the kernel required is $K(\underline{x}, \underline{y}) = (1+\underline{xy})^4$. Data in this higher dimension space will have a separator.

4. Consider the XOR data (located at $(\pm 1, \pm 1)$). Express the dual SVM problem and show that a separator can be found using

   $-K(x, y) = (1 + x.y)^2$
   $-K(x, y) = exp(-||x - y||^2)$.

   For each, determine the regions of (x_1, x_2) space where points will be classified as positive or negative. Given that each produces a distinct separator, how might you decide which of the two was preferred?

   The XOR data is x = $\begin{bmatrix} +1 & +1 \\ +1 & -1 \\ -1 & +1 \\ -1 & -1 \end{bmatrix}$

   and y = $\begin{bmatrix} -1 \\ +1 \\ +1 \\ -1 \end{bmatrix}$

   For the first part, the kernel function as $K(x, y) = (1 + x.y)^2$, we can calculate the values of the kernel matrix.

$$\begin{aligned} K_{11} &= (1 + \begin{bmatrix} -1 \\ -1 \end{bmatrix} \begin{bmatrix} -1 & -1 \end{bmatrix}) \\ &= (1+2)^2 \\ &= 9 \end{aligned}$$

   Similarly, the complete K-matrix is : K = $\begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix}$

   On multiplying the above K matrix with the result vector y = $\begin{bmatrix} -1 \\ +1 \\ +1 \\ -1 \end{bmatrix}$,

   Where,

$$K'_{ij} = K_{ij} y^i y^j$$

$$K' = \begin{bmatrix} 9 & -1 & -1 & 1 \\ -1 & 9 & 1 & -1 \\ -1 & 1 & 9 & -1 \\ 1 & -1 & -1 & 9 \end{bmatrix}$$

Now, according to Dual SVM, and taking its derivative

$$\begin{aligned} L(\alpha) &= \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i y^i K_{ij} y^j \alpha_j \\ \nabla(L) &= 1 - K\alpha \\ \alpha &= \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix} \end{aligned}$$

We see that the values for α satisfy the conditions imposed on dual SVM equation. Hence we can say it is a valid solution.

Now, let us find $w = \sum_i \alpha_i y_i \phi(x_i)$, using the feature mapping function as $\phi(x_1, x_2) = (1, x_1, x_2, x_1 x_2, x_1^2, x_2^2)$ which is the following:

$$w = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

There is no bias and the weight is only present for the term $x_1 x2$.

Hence the classifier is :

$$y = w^T \phi(x) = -x_1 x2$$

The solution to second kernel function follow here :

The kernel matrix is as follows :

$$\begin{aligned} K_{11} &= e^{(-\left\| \begin{bmatrix} +1 \\ +1 \end{bmatrix} - \begin{bmatrix} +1 \\ +1 \end{bmatrix} \right\|)} && (1) \\ &= e^{(-\left\| \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\|)} && (2) \\ &= 1 && (3) \end{aligned}$$

Similarly the complete K Matrix obtained is K= $\begin{bmatrix} 1 & e^{-4} & e^{-4} & e^{-8} \\ e^{-4} & 1 & e^{-8} & e^{-4} \\ e^{-4} & e^{-8} & 1 & e^{-4} \\ e^{-8} & e^{-4} & e^{-4} & 1 \end{bmatrix}$

Now, we multiply this matrix with y where $\tilde{K_{ij}} = K_{ij}y^i y^j$:

$$K' = \begin{bmatrix} 1 & -e^{-4} & -e^{-4} & e^{-8} \\ -e^{-4} & 1 & e^{-8} & -e^{-4} \\ -e^{-4} & e^{-8} & 1 & -e^{-4} \\ e^{-8} & -e^{-4} & -e^{-4} & 1 \end{bmatrix}$$

$$\begin{aligned} L(\alpha) &= \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i y^i K_{ij} y^j \alpha_j \\ \nabla(L) &= 1^T\alpha - \tilde{K}\alpha \\ \alpha &= \begin{bmatrix} 1.0376 & 1.0376 & 1.0376 & 1.0376 \end{bmatrix} \end{aligned}$$

We see that the values for α satisfy the conditions imposed on dual SVM equation. Hence we can say it is a valid solution. All the values of α are the same. Using these values, we will now compute the values of w and b as follows :

$$w = \sum_i \alpha_i y_i \phi(x_i)$$

The feature map in this case is - $\phi(x_1, x_2) = (e^{-x^2}, xe^{-x^2}, \frac{x^2 e^{-x^2}}{\sqrt{2}}, \frac{x^3 e^{-x^2}}{\sqrt{6}}, \dots)$.

Above feature map is used to obtained the w, which in turn helps in obtaining the classifier.

Both the kernel gives us separators in higher dimensional space, however the polynomial kernel $K(\underline{x}, \underline{y}) = (1 + \underline{x}.\underline{y})^2$ is preferred kernel. It finds a simple linear separator and is computationally less expensive as compared to the RBF kernel.