Kernelized SVM

25 October 2021 19:38

I had a doubt earlier :

Once you solve the Dual Problem, you will get your support vectors (those having non-zero α_i). After getting your support vectors, you can derive your ω^*.

$$\omega^* = \sum_{i=1}^{m} \alpha_i y_i x_i$$

Doubt...

[This raised a question in my mind, if we used a kernel function to get the α's, then we won't be able to use this formula because we won't have the x_i's in the higher dimensions.]

On going through 'Hands On Machine Learning with Scikit-Learn, Keras and TensorFlow', I came to know that the calculating of ω and b for Kernalized SVM is very similar to that of 'Simple' SVM.

Given the α's from solving the Dual Problem :

$$\boxed{\omega = \sum_{i=1}^{m} \alpha_i y_i \phi(x^{(i)})}$$

where $\phi(x^{(i)})$ is the transformed training example $x^{(i)}$.

However, we do not physically have $\phi(x^{(i)})$ because all a Kernel gives us is transformed inner-products $\phi(x^{(i)}).\phi(x^{(j)})$ to plug into the Dual Problem. :(

We actually don't need ω and b to get our hypothesis function, $\omega.x + b$ working !

$$h_{\omega,b}(\phi(x^{(j)})) = \omega.\phi(x^{(j)}) + b$$

$$h_{\omega,b}(\phi(x^{(j)})) = \left[\alpha_i y_i \phi(x^{(i)}) \right] . \phi(x^{(j)}) + b$$

$$h_{\omega,b}(\phi(x^{(j)})) = \left[\sum_{i=1}^{m} \alpha_i y_i \, \phi(x^{(i)})\right] \cdot \phi(x^{(j)}) + b$$

$$h_{\omega,b}(\phi(x^{(j)})) = \left[\sum_{i=1}^{m} \alpha_i y_i \left(\phi(x^{(i)}) \cdot \phi(x^{(j)})\right)\right] + b$$

$$\boxed{h_{\omega,b}(\phi(x^{(j)})) = \sum_{i=1}^{m} \alpha_i y_i K(x^{(i)}, x^{(j)}) + b \quad \text{where } \alpha_i > 0}$$

Note that since $\alpha_i \neq 0$ only for SVs, thus we will need the transformed inner products of $x^{(j)}$ only with the SVs.

$$b = \frac{1}{m} \sum_{i=1}^{m} \left(y_i - \omega \cdot \phi(x^{(i)})\right) = \frac{1}{m} \sum_{i=1}^{m} \left[y_i - \left(\sum_{j=1}^{m} \alpha_j y_j \phi(x^{(j)})\right) \cdot \phi(x^{(i)})\right]$$

$$b = \frac{1}{m} \sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{m} \alpha_j y_j \, \phi(x^{(j)}) \cdot \phi(x^{(i)})\right]$$

$$\boxed{b = \frac{1}{m} \sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{m} \alpha_j y_j K(x^{(i)}, x^{(j)})\right]}$$

Finally our hyperplane becomes:

$$\boxed{h_{\omega,b}(\phi(x^{(j)})) = \sum_{i=1}^{m} \alpha_i y_i K(x^{(i)}, x^{(j)}) + \frac{1}{m} \sum_{i=1}^{m} \left[y_i - \sum_{j=1}^{m} \alpha_j y_j K(x^{(i)}, x^{(j)})\right]}$$