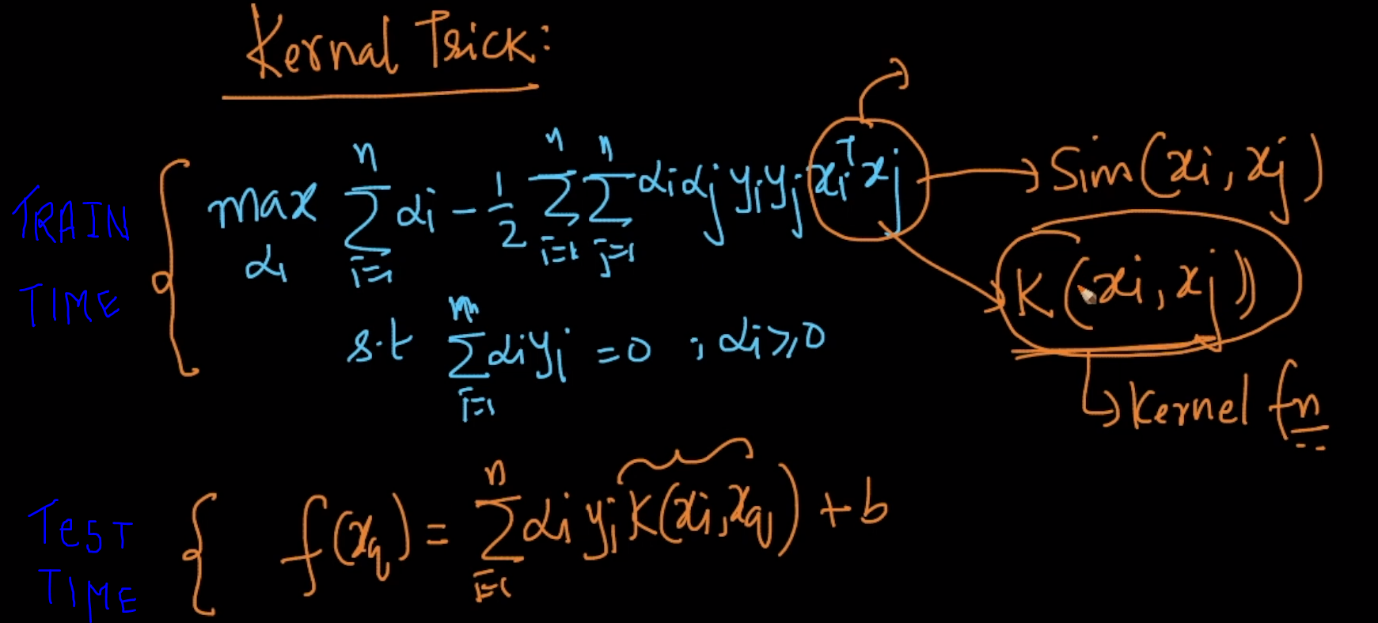
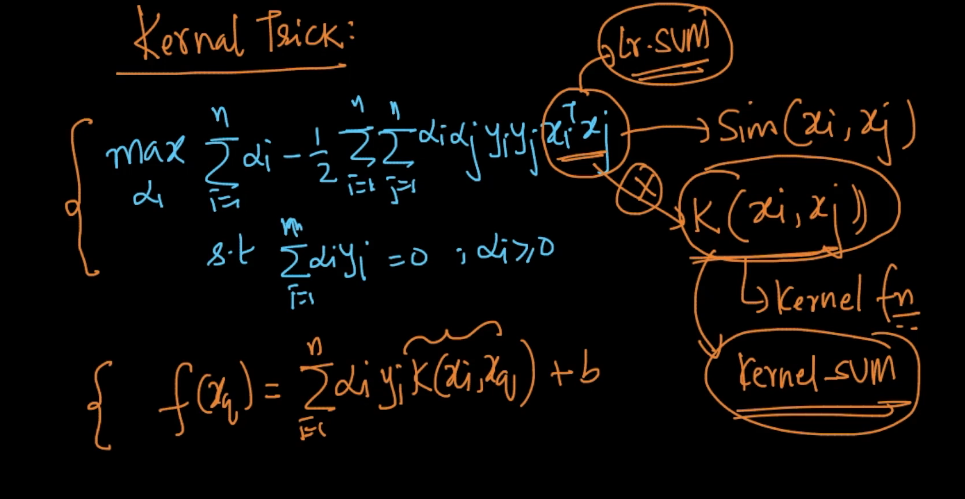
**kernel trick**

As shown in below figures in train time equation we can use any similarity function sim(Xi,Xj) instead of XTi Xj, one class of similarity function is kernel function.

We apply same kernel function in test/run time equation as well.



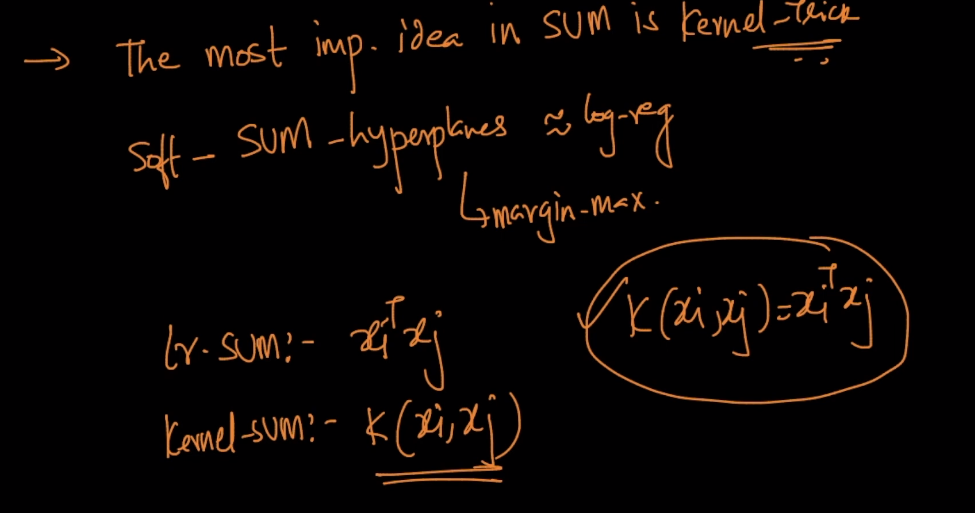
If we don’t use kernel function then it is linear SVM and if we replace XiXj with kernel function then it is kernel SVM.

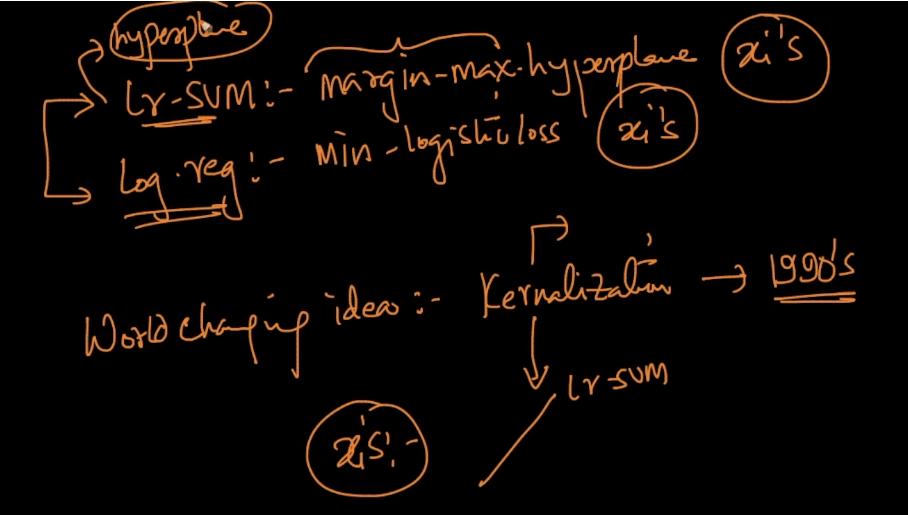


Soft\_SVM\_hyperplane is approx. same as logistic regression because we find plane in both to separate data.

In linear SVM we find margin-maximal hyperplane in Xi’s space

And in kernel SVM we find min logistic loss in Xi’s space.





if given data is in pattern as show below then both linear SVM and logistic regression fail to apply any hyperplane to separate this data.

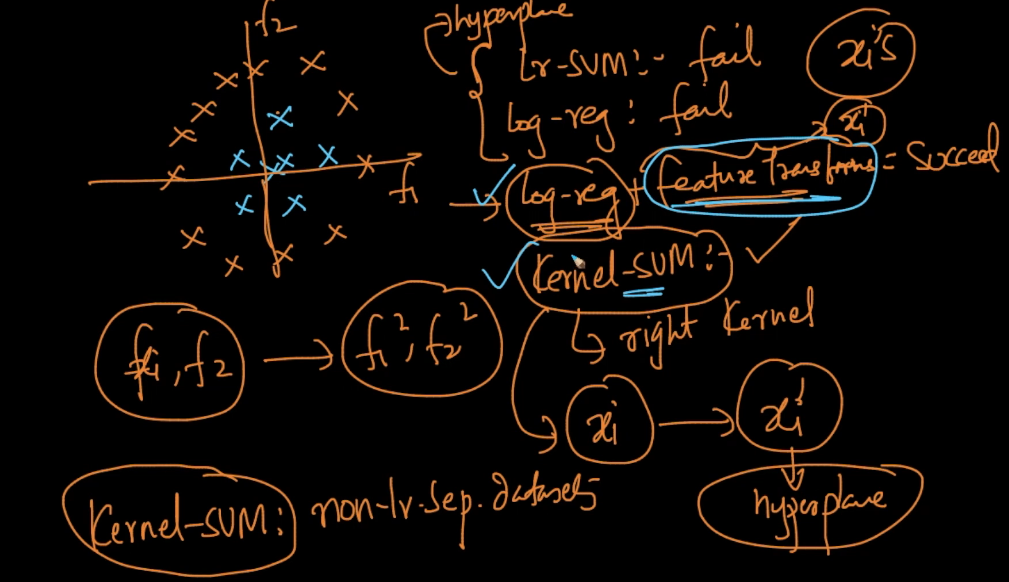
But logistic regression + feature transformation succed to separate this data.

But for this we have to feature transformation

To overcome this Kernel-SVM comes by choosing right kernel we can separate this type of data as well.

Kernel works like : It is not possible to find a hyperplane or a linear decision boundary for some classification problems. If we project the data in to a higher dimension from the original space, we may get a hyperplane in the projected dimension that helps to classify the data. Kernel helps to find a hyperplane in the higher dimensional space . Thus, In Support Vector Machine, any linear model can be turned into a non-linear model by applying the kernel trick.

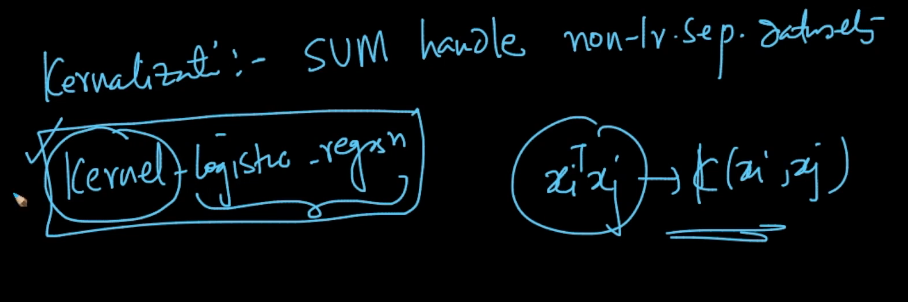
That means for a given pair of vectors (in a lower-dimensional feature space) and a transformation into a higher-dimensional space, there exists a function (The Kernel Function) which can compute the dot product in the higher-dimensional space without explicitly transforming the vectors into the higher-dimensional space first, unlike previously where we have to do feature transformation manually using domain knowledge.  
  
For datasets with a lot of features, it becomes next to impossible to try out all the interesting transformations. Hence using kernel trick we are saved !



Kernalization SVM hadle non-linear Separable datasets.

In literature there is also kernel logistic regression which replace XTiXj with kernel function.

And in general most supervised algorithms have a kernelized version as well.



Comments :

