Neural Networks for Signal Processing-I

Prof. Shayan Srinivasa Garani

Department of Electronic System Engineering

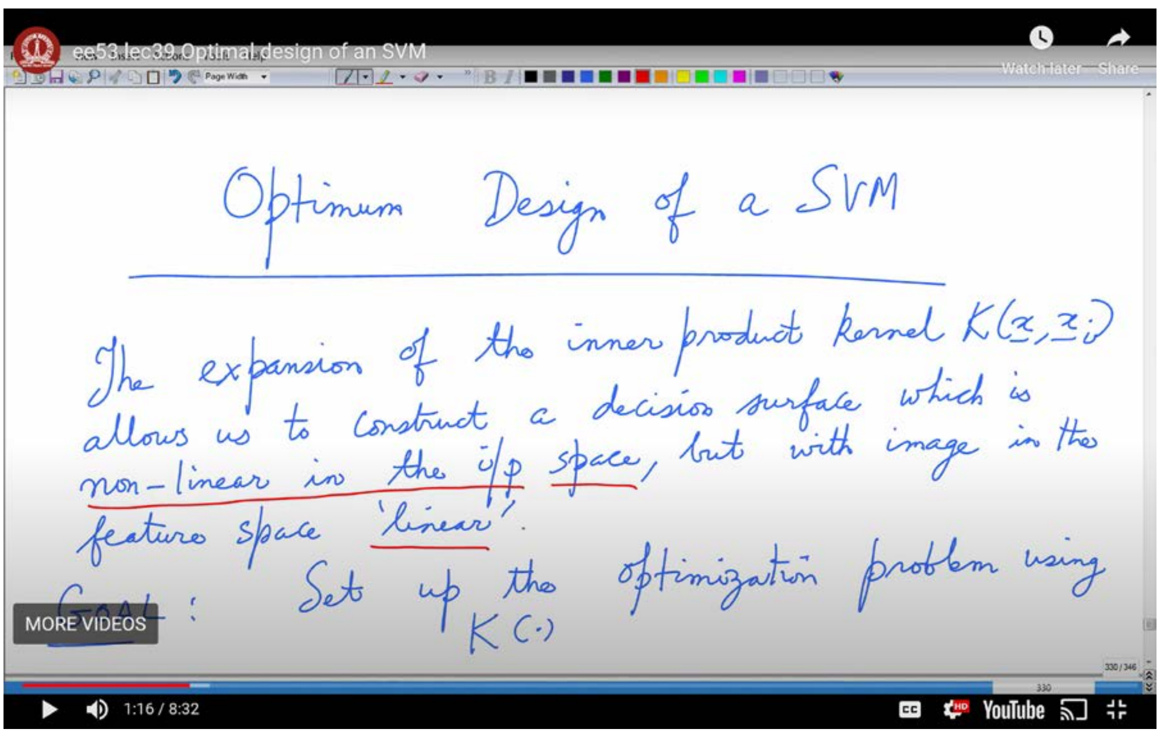
Indian Institute of Science, Bengaluru

Lecture – 39

Optimal Design of an SVM

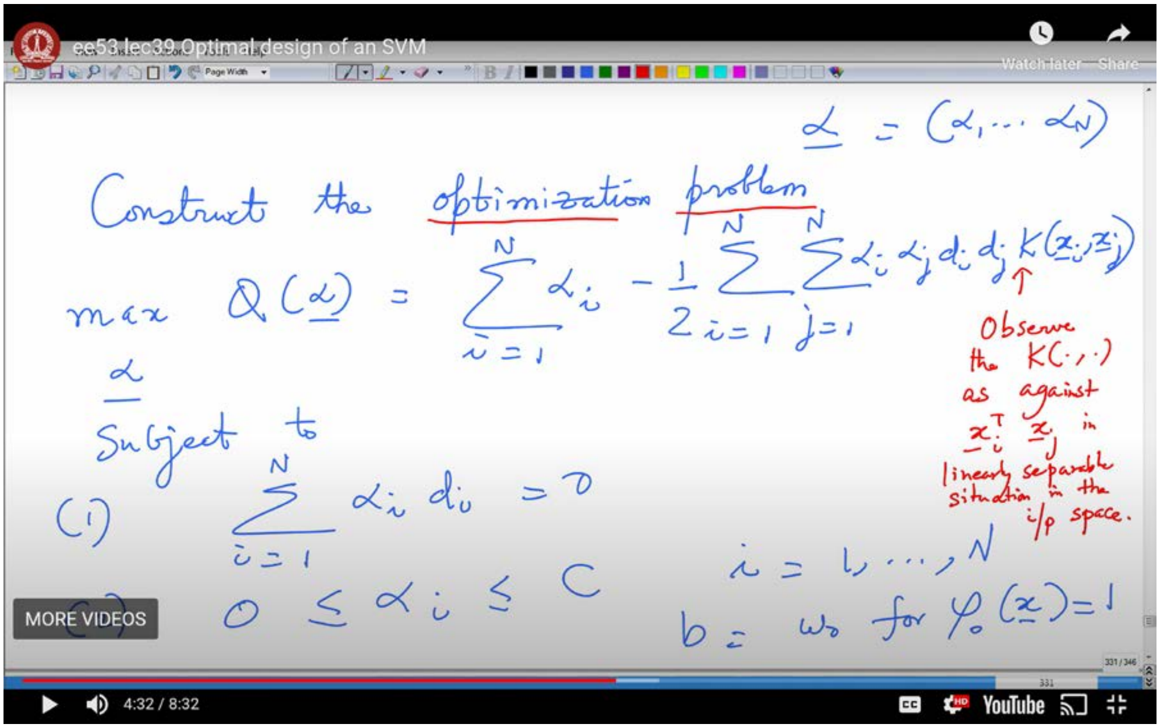
Having explored the concept of inner product kernels, let’s now shift our focus to the optimal design of Support Vector Machines (SVMs) using these kernels. The expansion of the inner product kernel $k(x, x_i)$ enables us to construct a decision surface that is non-linear in the input space but linear in the feature space. This is a key takeaway: while the relationship in the input space is non-linear, it becomes linear in the transformed feature space. Our objective is to formulate the optimization problem utilizing this kernel function. Let’s delve into the details.

(Refer Slide Time: 01:16)



We can set up the optimization problem as follows: We aim to maximize the function Q(α), where α is a vector composed of the Lagrange multipliers $\alpha_1, \alpha_2, \dots, \alpha_n$. The goal is to maximize Q(α) with respect to α, subject to specific constraints.

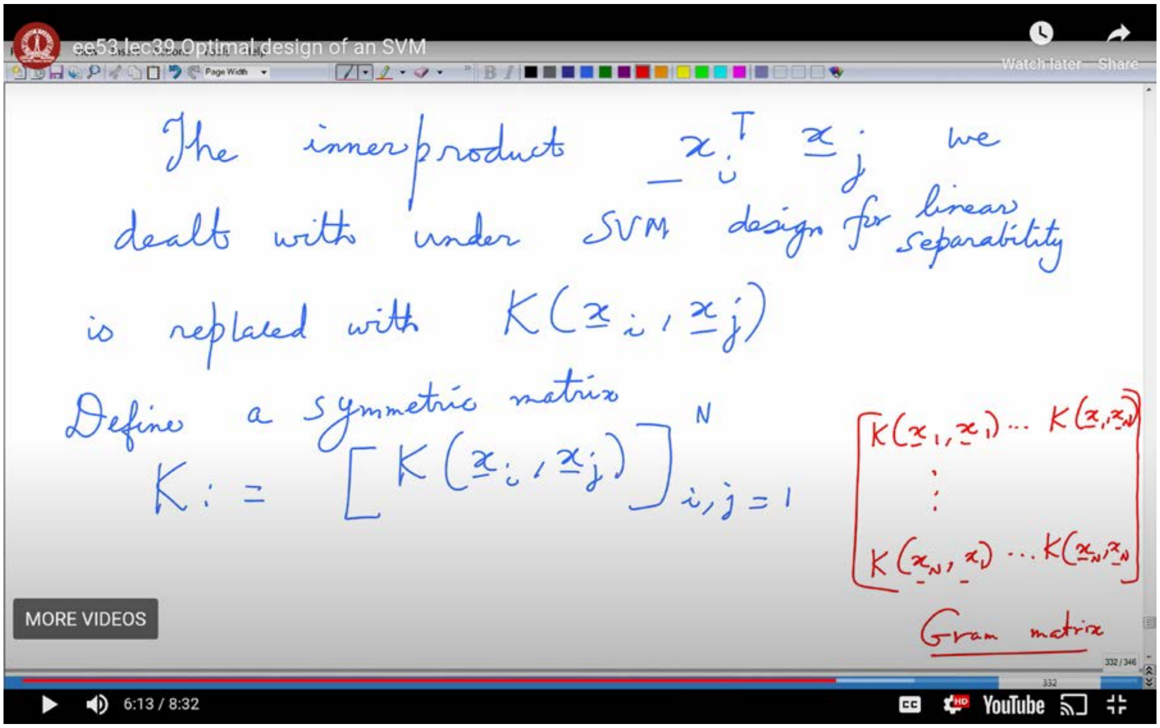
(Refer Slide Time: 04:32)



In the context of linear separability, the dual problem involves summing over the Lagrange multipliers α_i, and subtracting a term that includes the Lagrange multipliers, the desired responses, and the kernel acting on the data points. This formulation mirrors the approach used in SVM design for linearly separable patterns. In the input space, if the patterns are linearly separable, we could express this in terms of $x_i^T x_j$ as we did earlier. However, notice that in this non-linear space, the kernel $k(x_i, x_j)$ is now applied, as we have transformed the inputs non-linearly into the feature space. Our task then is to find a linear classifier in this feature space to achieve pattern separability.

It's important to observe that in the linearly separable scenario in the input space, we dealt with $x_i^T x_j$, but now we focus on the kernel $k(x_i, x_j)$ in the non-linear feature space.

(Refer Slide Time: 06:13)



Now, let’s consider the constraints for this optimization problem. We have the condition:

$$\sum_{i=1}^{n} \alpha_i d_i = 0$$

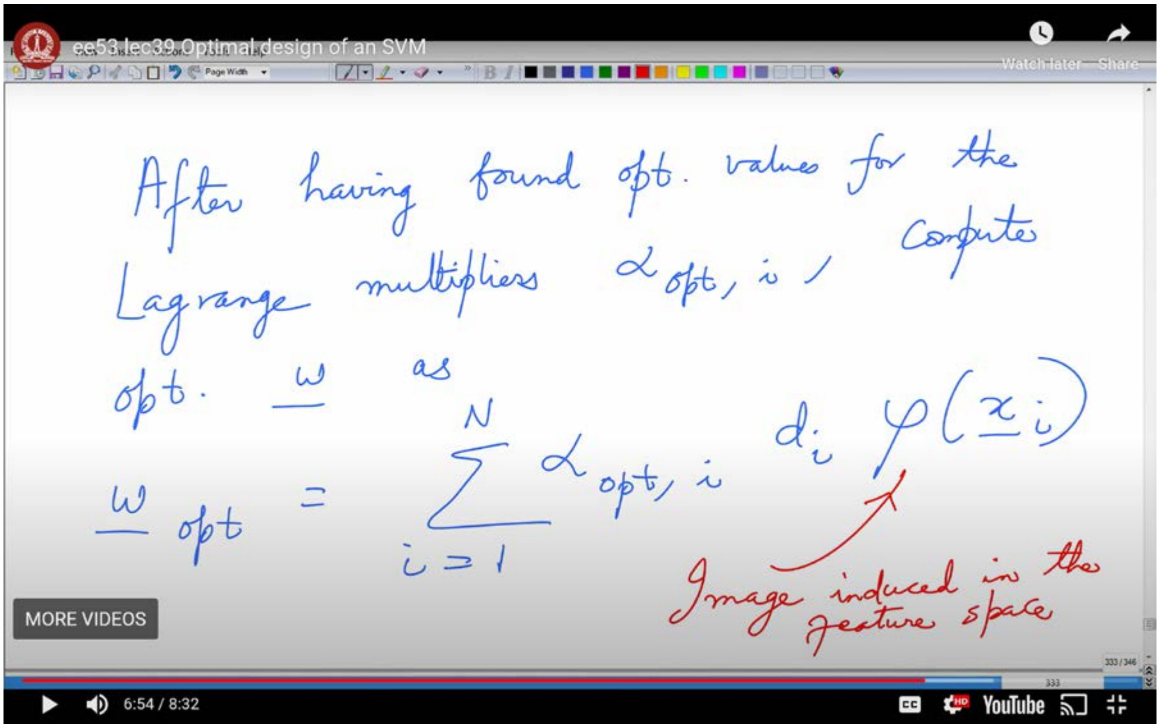
Additionally, each α_i is bounded between 0 and a constant C, where C is a fixed value. These constraints must hold for all i from 1 to n. The bias term is also incorporated when solving for the kernel in terms of the non-linear mapping.

This formulation is straightforward yet powerful. The core idea is to "lift" the vector from the input space to the feature space, where we then work with a linear classifier.

The inner product kernel $k(x_i, x_j)$ can be computed for all pairs i and j from 1 to n. You can easily verify that this results in a symmetric matrix, which is an essential property. This matrix is often referred to as the Gram matrix, and it represents the inner product of the vectors x_i and x_j in the feature space.

Visualize this matrix as a grid where each entry corresponds to $k(x_i, x_j)$. The matrix remains symmetric, reflecting the underlying structure of the kernel in the feature space.

(Refer Slide Time: 06:54)



Having determined the optimal values for the Lagrange multipliers, we proceed by taking the derivatives with respect to the parameters in the Lagrangian and solving for these optimal multipliers. Once we have the optimal Lagrange multipliers, we can compute the optimal weight vector w. This is done using the formula:

$$w = \sum_{i=1}^{N} \alpha_i^* d_i \phi(x_i)$$

where α_i^* represents the optimal Lagrange multipliers, d_i is the desired response, and $\varphi(x_i)$ is the image of the input vector x_i in the feature space. This process allows us to calculate the optimal weight vector straightforwardly.

With the optimal weight vector in hand, we can solve the pattern separability problem, thus completing the design of the Support Vector Machine (SVM) for cases where patterns are linearly separable in the feature space.

When dealing with non-separable patterns in the input space, the approach involves lifting the vector non-linearly into the feature space, setting up the optimization problem accordingly, and solving it to obtain the optimal Lagrange multipliers. From these, we can determine the hyperplane. This methodology effectively handles cases of non-linear separability.

In scenarios where patterns are not linearly separable and may be misclassified, we adapt the approach by incorporating conditions to minimize the average error due to misclassification. We then formulate the cost functional based on this risk computation and proceed with the optimization in a manner similar to the earlier steps. This concludes our discussion on the design of SVMs for non-linear separable patterns in the input space.