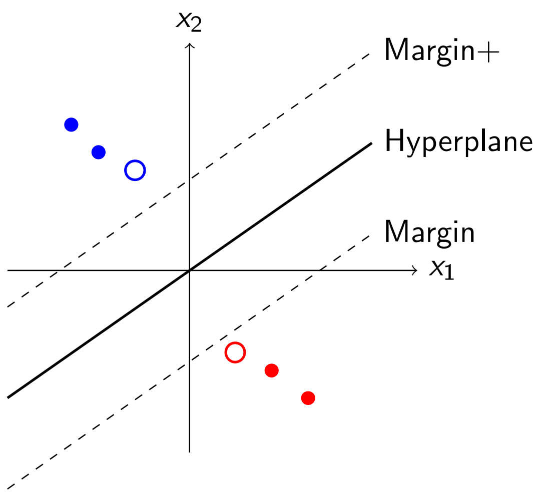
Mathematical Foundation of Support Vector Machines (SVM)

Dr. Shailesh Sivan

What is an SVM?

- A supervised learning algorithm for classification.
- Goal: Find the optimal hyperplane that separates classes.
- Based on:
  - Maximum margin
  - Convex optimization
  - Kernel methods

Hyperplane and Margin (Geometric View)



Hard-Margin SVM: Optimization

$$\min_{w,b} \quad \frac{1}{2}\|w\|^2$$

$$\text{subject to} \quad y_i(w^\top x_i + b) \geq 1$$

- Maximum margin minimize norm of w
- Only holds when data is linearly separable

Dual Formulation

$$\max_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^\top x_j$$

$$\text{subject to} \quad \sum_{i=1}^{n} \alpha_i y_i = 0, \quad \alpha_i \geq 0$$

- Dual uses dot products $\rightarrow$ allows use of kernels
- Solution: $w = \sum_i \alpha_i y_i x_i$

Support Vectors

- Data points for which $\alpha_i > 0$
- They lie closest to the decision boundary
- Define the hyperplane
- Others have no impact on the solution

Soft-Margin SVM

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n} \xi_i$$

$$\text{subject to} \quad y_i(w^\top x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

- Allows misclassification using slack ξ_i
- C: Penalty for misclassification

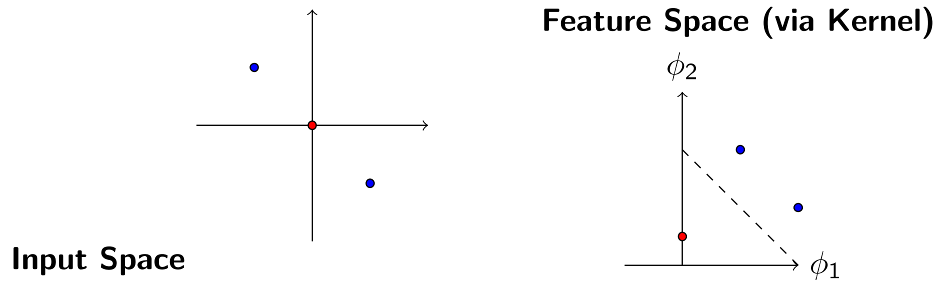
Kernel Trick: Non-Linear Boundaries

- Map data to high-dimensional space:

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D, \quad D \gg d$$

- Kernel function: $K(x_i, x_j) = \phi(x_i)^\top \phi(x_j)$
- Avoids explicit computation of ϕ

Kernel Transformation Illustration



Common Kernel Functions

- Linear: $K(x, x') = x^\top x'$
- Polynomial: $K(x, x') = (x^\top x' + c)^d$
- RBF (Gaussian): $K(x, x') = \exp(-\gamma \|x - x'\|^2)$
- Sigmoid: $K(x, x') = \tanh(\kappa x^\top x' + c)$

Summary

- SVM finds a maximum-margin hyperplane
- Uses convex optimization and support vectors
- Can handle non-linear data using the kernel trick
- Powerful for both classification and regression