

# Support Vector Machines and Principal Component Analysis

A Short Report

## 1. Introduction

Machine learning often involves finding meaningful patterns in high-dimensional data. Two widely used techniques are **Support Vector Machines (SVM)** for classification and regression, and **Principal Component Analysis (PCA)** for dimensionality reduction. Both rely on mathematical optimization and linear algebra, but serve different purposes.

## 2. Support Vector Machines (SVM)

### 2.1. What is SVM?

Support Vector Machines are supervised learning algorithms used for classification and regression. The core idea is to find an optimal hyperplane that separates data points of different classes with the *maximum margin*.

### 2.2. Mathematical Formulation

Given training data  $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$  where  $\mathbf{x}_i \in \mathbb{R}^d$  and  $y_i \in \{-1, +1\}$ , the goal is to find a hyperplane

$$\mathbf{w}^T \mathbf{x} + b = 0$$

that best separates the two classes. The optimization problem is:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{subject to } y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \forall i$$

This is a convex quadratic optimization problem. The data points that lie closest to the decision boundary are called **support vectors**.

For non-linearly separable data, SVM uses a **kernel function**  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j)$  to implicitly map inputs into a higher-dimensional space.

### 2.3. Advantages and Pitfalls

**Advantages:**

- Works well for high-dimensional data.
- Robust to overfitting when using appropriate regularization.

- The kernel trick enables handling of non-linear data.

#### **Pitfalls:**

- Choosing the right kernel and parameters (e.g.,  $C$ ,  $\gamma$ ) can be difficult.
- Computationally expensive for large datasets.

## **3. Principal Component Analysis (PCA)**

### **3.1. What is PCA?**

PCA is an unsupervised learning method used for dimensionality reduction. It transforms correlated variables into a set of linearly uncorrelated variables called **principal components**, ordered by the amount of variance they explain in the data.

### **3.2. Mathematical Formulation**

Given a dataset  $X \in \mathbb{R}^{n \times d}$ , PCA centers the data by subtracting the mean and computes the covariance matrix:

$$\Sigma = \frac{1}{n} X^T X$$

The principal components are obtained by solving the eigenvalue problem:

$$\Sigma \mathbf{v}_i = \lambda_i \mathbf{v}_i$$

where  $\mathbf{v}_i$  are eigenvectors (principal axes) and  $\lambda_i$  are eigenvalues representing variance captured by each component. The first  $k$  eigenvectors with the largest eigenvalues form the reduced feature space:

$$Z = XW_k, \quad W_k = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$$

### **3.3. Advantages and Pitfalls**

#### **Advantages:**

- Reduces dimensionality while preserving variance.
- Removes noise and multicollinearity.
- Improves visualization and computational efficiency.

#### **Pitfalls:**

- Principal components may be hard to interpret.
- Sensitive to data scaling and outliers.
- Assumes linear relationships.

## 4. Conclusion

SVM and PCA are foundational techniques in machine learning. While SVM focuses on building powerful classifiers with optimal decision boundaries, PCA aims to simplify data by reducing redundancy. Together, they often form a powerful preprocessing and modeling pipeline — with PCA used for feature extraction and SVM for final classification.