Support Vector Machines and Principal Component Analysis

Theoretical Analysis of Support Vector Machines and Principal Component Analysis

Atharv Bhatt

October 24, 2025

1 Introduction and Project Scope

This report presents a theoretical analysis of two fundamental machine learning techniques: *Support Vector Machines (SVM)* and *Principal Component Analysis (PCA)*. The objective is to explain their mathematical foundations, conceptual frameworks, and working principles. These techniques are widely used in modern machine learning pipelines for classification and dimensionality reduction tasks.

2 Support Vector Machines (SVM)

2.1 What is SVM and Why is it Used?

Support Vector Machines (SVM) are supervised learning algorithms primarily used for binary classification, with extensions to regression problems. The central idea of SVM is to find an optimal decision boundary, known as a *hyperplane*, that maximally separates data points belonging to different classes.

The data points closest to the hyperplane are called *support vectors*. These points play a critical role in defining the position and orientation of the decision boundary. SVMs are particularly effective in high-dimensional spaces and when the data is not linearly separable. By relying only on the support vectors, SVMs are memory-efficient and often exhibit strong generalization performance.

2.2 Model Framework and Mathematical Formulation

2.2.1 Hard Margin Classification

For linearly separable data, the objective is to find a hyperplane

$$w \cdot x + b = 0$$

that maximizes the margin between the two classes. This is formulated as the following optimization problem:

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

subject to:

$$y_i(w \cdot x_i + b) \geq 1, \quad \forall i$$

2.2.2 Soft Margin Classification

When the data is not perfectly separable, slack variables ζ_i are introduced:

$$\min_{w,b,\zeta} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\zeta_i$$

subject to:

$$y_i(w \cdot x_i + b) \geq 1 - \zeta_i, \quad \zeta_i \geq 0$$

The regularization parameter C controls the trade-off between maximizing the margin and penalizing misclassification errors.

2.2.3 The Kernel Trick

To handle non-linear decision boundaries, SVM employs the *kernel trick*, which implicitly maps the input data into a higher-dimensional feature space:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Common kernel functions include:

- Polynomial Kernel:

$$K(x_i, x_j) = (\gamma x_i \cdot x_j + r)^d$$

- Radial Basis Function (RBF) Kernel:

$$K(x_i, x_j) = \exp(-\gamma\|x_i - x_j\|^2)$$

2.2.4 Dual Optimization and Decision Function

The dual optimization problem is given by:

$$\max_{\alpha} \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

subject to:

$$\sum_{i=1}^{m} \alpha_i y_i = 0, \quad 0 \leq \alpha_i \leq C$$

The final decision function is:

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i K(x_i, x) + b$$

2.3 Advantages and Pitfalls

Advantages

- Effective in high-dimensional spaces
- Strong theoretical guarantees due to convex optimization
- Memory efficient, relying only on support vectors
- Flexible through kernel selection

Pitfalls

- Computationally expensive for large datasets
- Sensitive to kernel choice and hyperparameters
- Does not naturally provide probabilistic outputs

3 Principal Component Analysis (PCA)

3.1 What is PCA and Why is it Used?

Principal Component Analysis (PCA) is an unsupervised dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while preserving as much variance as possible. It achieves this by identifying orthogonal directions called *principal components*.

3.2 Mathematical Framework

3.2.1 Data Standardization

Prior to applying PCA, the data is mean-centered and typically standardized. This is crucial because PCA is sensitive to the scale of the features.

3.2.2 Eigen Decomposition

The steps involved in PCA are:

1. Compute the covariance matrix C.
2. Solve the eigenvalue problem:
$$Cv = \lambda v$$
3. Sort eigenvectors in descending order of eigenvalues.
4. Project the data onto the selected principal components.

3.3 Advantages and Pitfalls

Advantages

- Reduces computational complexity
- Removes correlated features
- Helps mitigate overfitting

Pitfalls

- Reduced interpretability
- Information loss is unavoidable
- Sensitive to feature scaling

4 Conclusion

Support Vector Machines and Principal Component Analysis are essential techniques in machine learning serving complementary roles. SVM provides a robust supervised learning framework for classification using margin maximization and kernel methods, while PCA is a powerful preprocessing tool for dimensionality reduction and noise removal. Together, they are widely used to build efficient and scalable machine learning systems.

References

1. StatQuest with Josh Starmer, *Support Vector Machines*, YouTube.
2. ritvikmath, *Support Vector Machines and PCA*, YouTube.
3. GeeksforGeeks, *Support Vector Machine Algorithm.*