3.2 - Support Vector Machines

Giovanni Della Lunga
giovanni.dellalunga@unibo.it

Bologna - February-April, 2025

Introduction

Linear Separation

- **Linearly Separable Data points**: Data points can be said to be linearly separable if a separating boundary/hyperplane can easily be drawn showing distinctively the different class groups.
- Linear separable data points mostly require linear machine learning classifiers such as Logistic regression for example.

../5-pictures/chapter-4-4_pic_

Linear Separation

../5-pictures/chapter-4-4_pic_1.pn

Perceptron

Perceptron Overview

- Introduced by Frank Rosenblatt in 1958.
- Used for binary classification problems.
- A single-layer neural network with step activation.

Perceptron Structure

- Inputs with associated weights.
- Weighted sum plus bias.
- Step function activation.

Mathematical Formulation

$$y = \begin{cases} 1, & \text{if } \sum w_i x_i + b \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Perceptron Learning Rule

- Initialize weights and bias.
- Update rule: $w_i \leftarrow w_i + \eta(y - \hat{y})x_i$
- Bias update: $b \leftarrow b + \eta(y - \hat{y})$

Perceptron Convergence Theorem

- The perceptron algorithm converges if the data is linearly separable.
- The number of updates is bounded by the margin of separation.

Support Vector Machines (SVMs)

What is an SVM?

- A powerful classification model.
- Finds the optimal separating hyperplane.
- Maximizes the margin between classes.

Mathematical Definition

$$\min \frac{1}{2}||w||^2 \quad \text{subject to } y_i(w \cdot x_i + b) \geq 1$$

Support Vectors and Margins

- Support vectors are closest to the decision boundary.
- The margin is the distance between these points and the hyperplane.

Hard vs Soft Margin SVM

- Hard Margin: No misclassification allowed.
- Soft Margin: Allows some misclassification to improve generalization.

Kernel Methods

The Kernel Trick

- Maps data to a higher-dimensional space.
- Avoids explicit computation of feature transformation.
- Enables SVMs to work with non-linearly separable data.

Common Kernel Functions

- Polynomial Kernel: $K(x, y) = (x \cdot y + c)^d$
- Radial Basis Function (RBF) Kernel: $K(x, y) = e^{-\gamma ||x-y||^2}$

Example: RBF Kernel in Python

Python Code:

```
from sklearn.svm import SVC
model = SVC(kernel='rbf', gamma=1)
model.fit(X_train, y_train)
```

Hyperparameter Tuning in SVMs

- C: Controls trade-off between margin width and misclassification.
- Gamma: Defines influence of a single training example.
- Kernel Choice: Linear, Polynomial, RBF.