Class 7

Support Vector Machine (SVM)

Organized By:
HSTU Machine Learning Enthusiasts

Prepared by: Amrito Das Tipu

Date: November 23, 2023

Outline

1. Introduction
2. Importance
3. Terminology
4. Types
5. Kernel
6. Kernel Trick
7. Advantages and disadvantages
8. Implementation
9. Conclusion

Previous classes

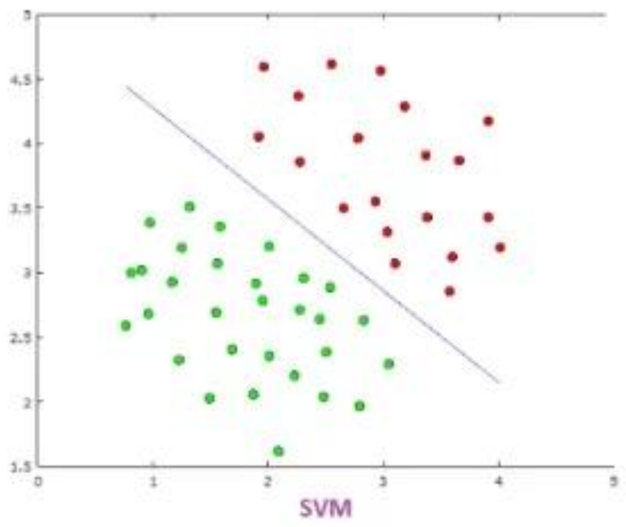
Class 1	• ML Research and Introduction, Anaconda, Jupyter notebook, Google Colab
Class 2	• Python Basics
Class 3	• Library Usage (Pandas, NumPy, Matplotlib)
Class 4	• Linear Regression
Class 5	• Data Preprocessing and Exploratory Data Analysis (EDA)
Class 6	• Logistic Regression

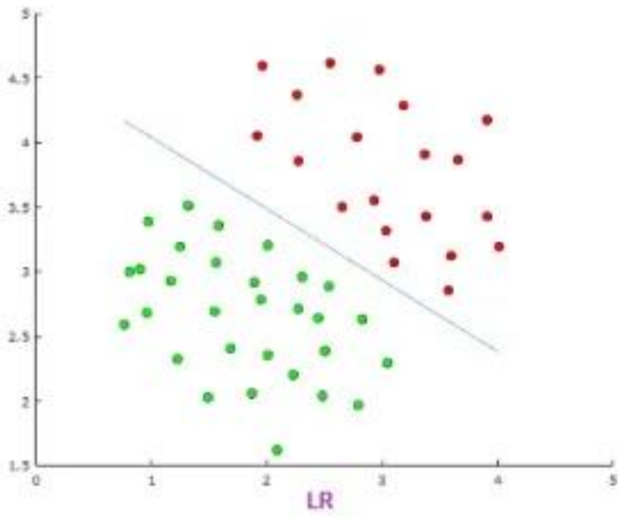
Introduction

- Developed at AT&T Bell Laboratories in 1990s
- Support Vector Machine or SVM
- One of the most robust prediction methods
- Based on statistical learning frameworks or VC theory
- A machine learning algorithm that uses supervised learning models to solve complex problems.
- Solves classification, and regression problems.

Introduction

Linear Regression	Logistic Regression	SVM
Regression tasks	Classification tasks	Both
Optimal decision boundary	Optimal decision boundary	Decision boundary with best margin
Linear data	Linear data	Linear and non-linear data
Vulnerable to overfitting	Vulnerable to overfitting	Lower chance of overfitting



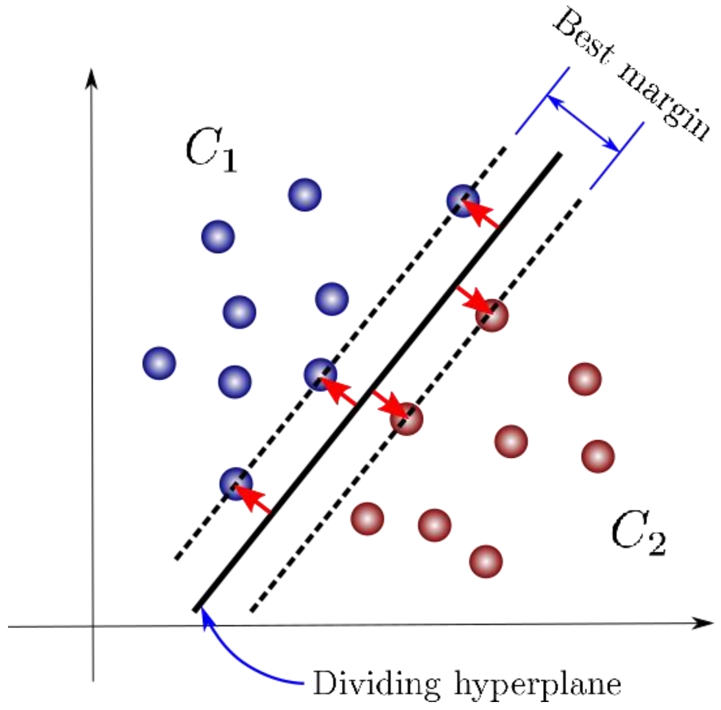


Importance

- It can handle both classification and regression on linear and non-linear data.
- Able to handle small but complex datasets.
- Real-world applications
  - Text and Hypertext Categorization
  - Handwriting Recognition
  - Speech Recognition
  - Image Classification
  - Security: Face detection, Intrusion detection
  - Bioinformatics: Protein classification and Cancer classification
  - Biomedical Science
  - Finance

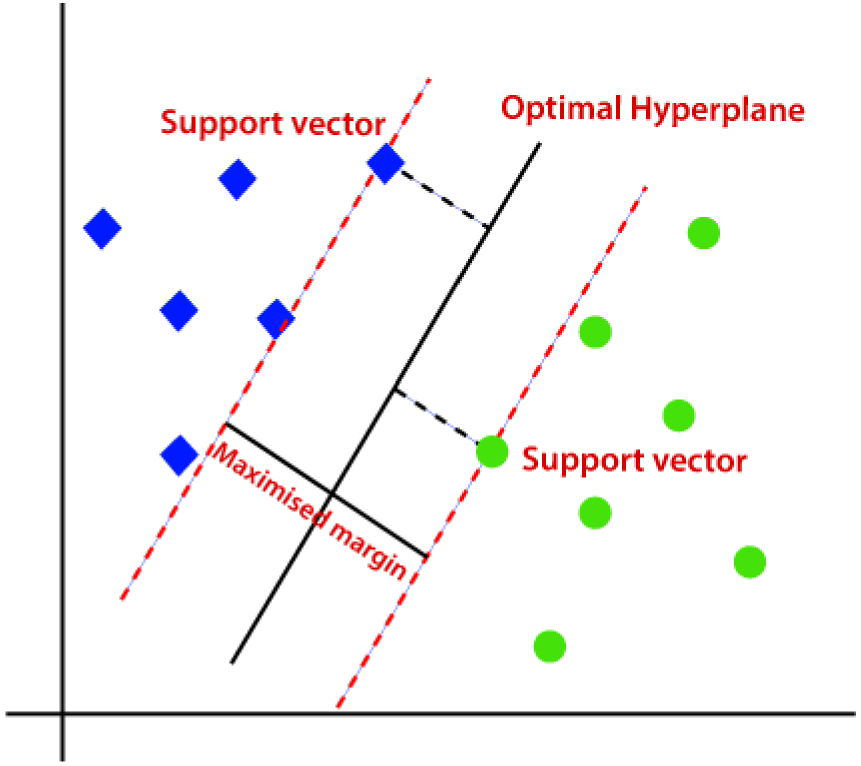
Terminology: Support Vectors

- Data points that are closest to the hyperplane.
- Define the decision boundary and the margin of separation.
- Critical elements of the training set.
- Deleting them will change the position of the hyperplane.



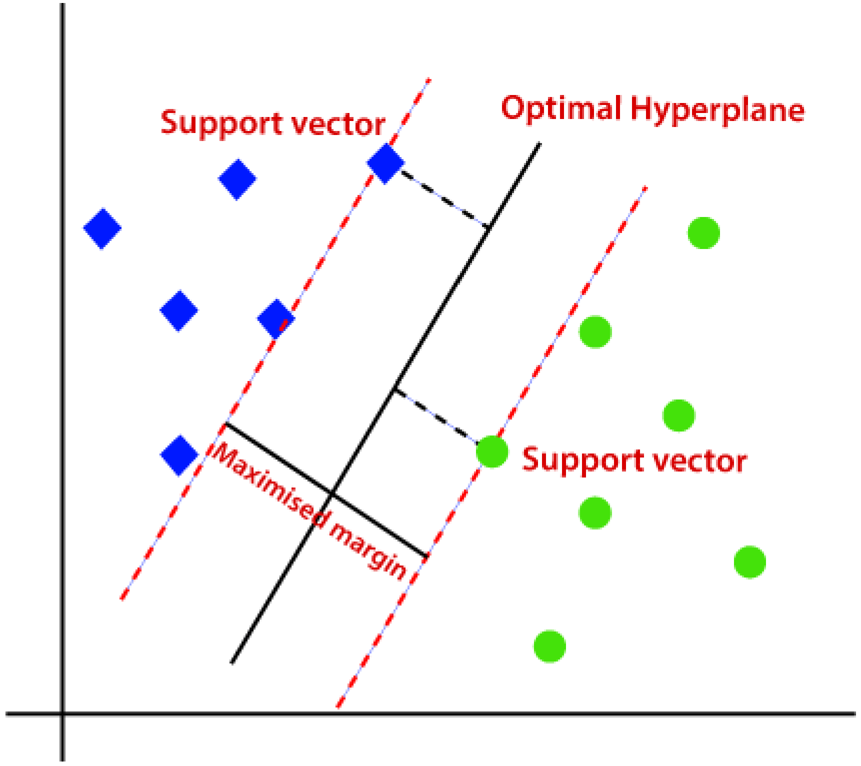
Terminology: Hyperplane

- A decision boundary that separates two classes of data.
- Best hyperplane is the one with the largest margin between the two classes.
- The objective of the SVM algorithm is to find a hyperplane that separates data points of one class from those of another class.
- For linear classifications, it will be a linear equation, $wx+b = 0$.



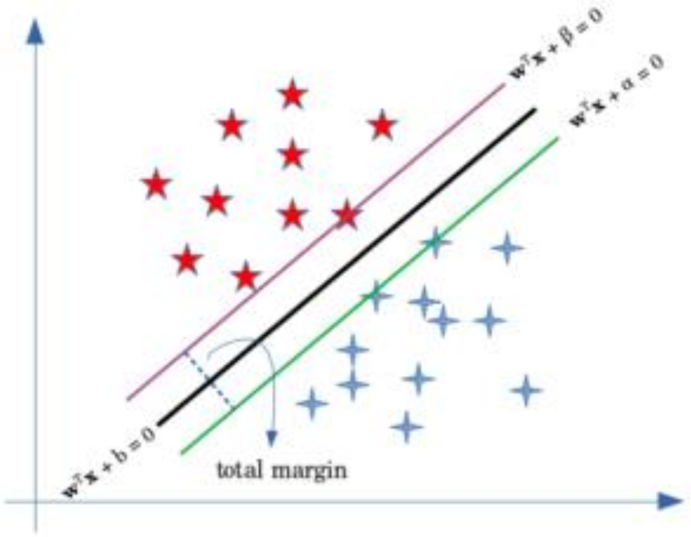
Terminology: Margin

- The distance between the support vector and hyperplane.
- A large margin is considered good.
  - Avoid random noise
  - Reduce overfitting
  - Reduce the number of potential classifiers
  - Reduce the possibility of generalization error

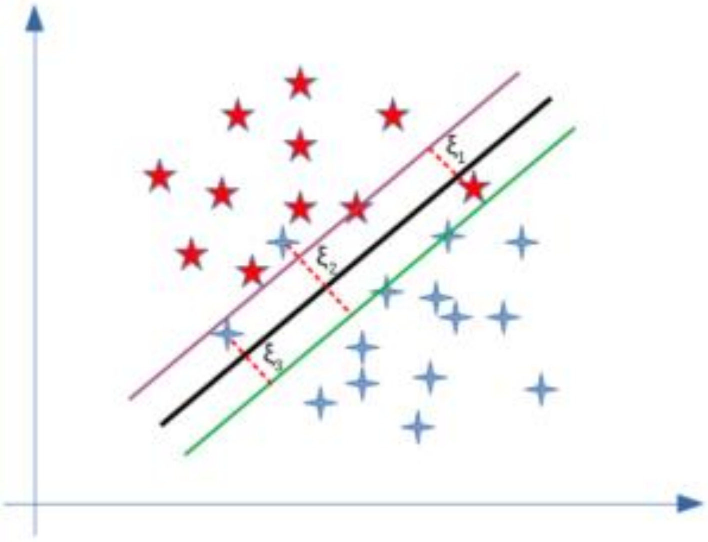


Terminology: Types of Margin

- Hard Margin
  - Separates data points of different categories without any misclassifications.
  - Use when data is linearly separable.

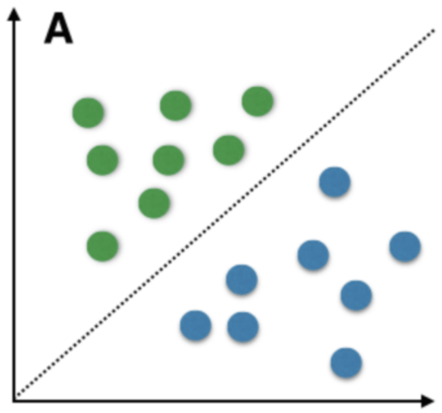


- Soft Margin
  - Allows some misclassifications for achieving better generality.
  - Each data point has a slack variable introduced in it.
  - When the data is not perfectly separable or contains outliers.

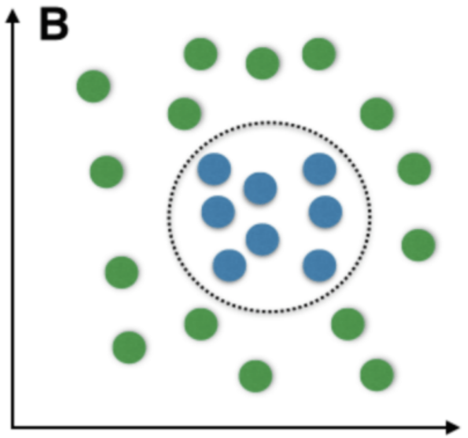


Types of SVM

- Linear SVM
  - When the data is perfectly linearly separable.
  - Data points can be classified into 2 classes by using a single straight line.

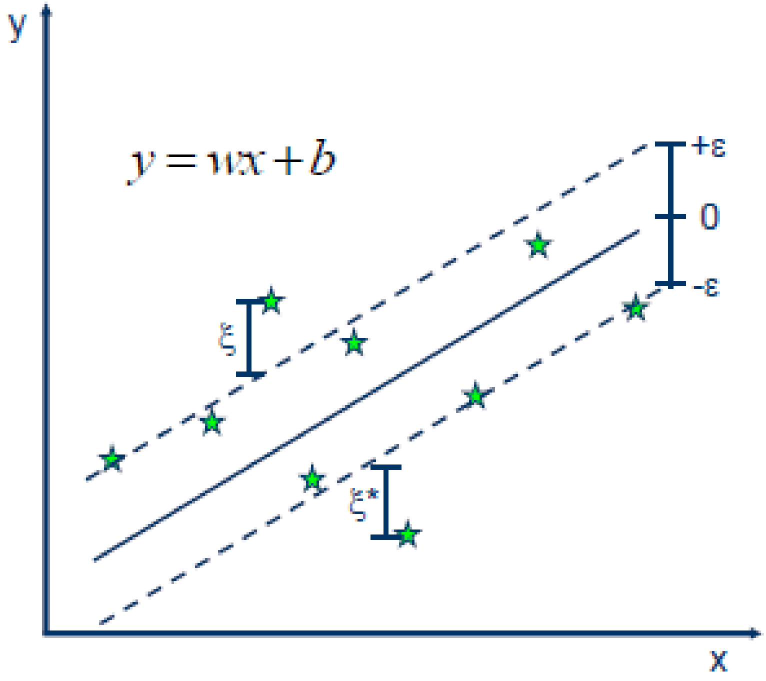


- Non-Linear SVM
  - When the data is not linearly separable.
  - Use kernel tricks to classify them.



Mathematical Formulation of SVMs

- Complex
- More details [here]
- Not necessary to know…



- Minimize:

$$\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\left(\xi_i + \xi_i^*\right)$$

- Constraints:

$$y_i - wx_i - b \le \varepsilon + \xi_i$$
$$wx_i + b - y_i \le \varepsilon + \xi_i^*$$
$$\xi_i, \xi_i^* \ge 0$$

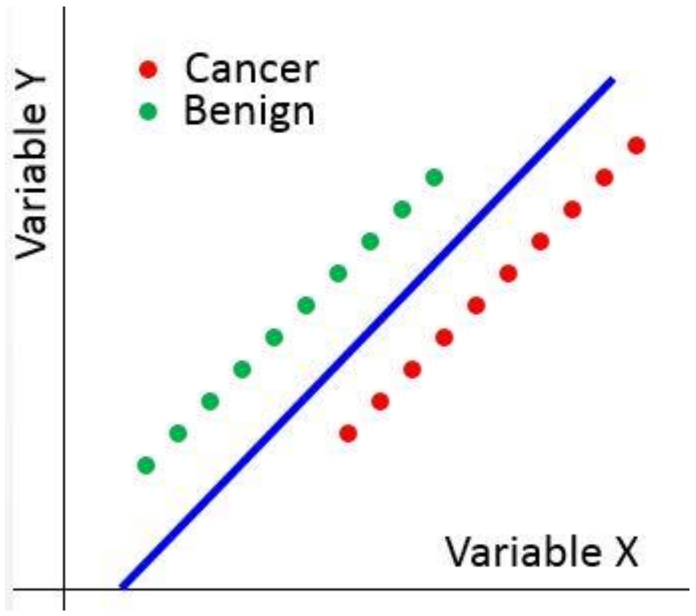
Kernel

- The mathematical function that maps the original input data points into high-dimensional feature spaces.
- Reason is to find the hyperplane when data points are not linearly separable in the original input space.
- Used to determine the decision boundary between different classes.
- Most common kernel functions are -
  - Linear
  - Polynomial
  - Radial Basis Function (RBF)
  - Sigmoid

Linear Kernel

- Learns a linear decision boundary in the original feature space.
- Used when data is linearly separable.

$$k(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1 \cdot \mathbf{x}_2$$



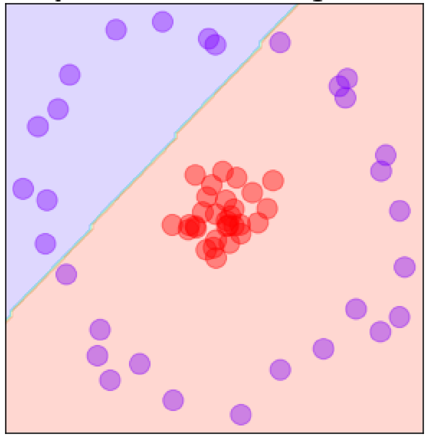
Polynomial Kernel

- Represents the similarity of vectors in a feature space over polynomials of the original variables.
- A polynomial kernel uses a polynomial function to map data into a higher-dimensional space.

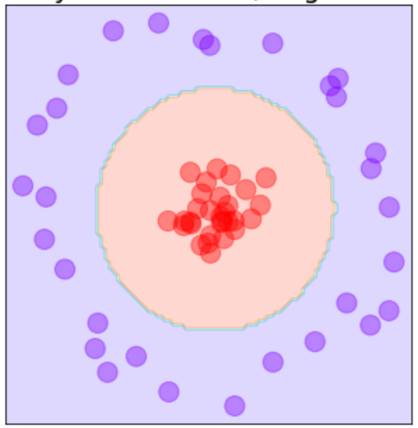
$$k(\mathbf{x}_1, \mathbf{x}_2) = (\gamma\, \mathbf{x}_1 \cdot \mathbf{x}_2 + c)^d$$

- Benefits are:
  - Detects both linear and nonlinear correlations in the data
  - Uses a non-linear function as kernel

Polynomial kernel, degree=1



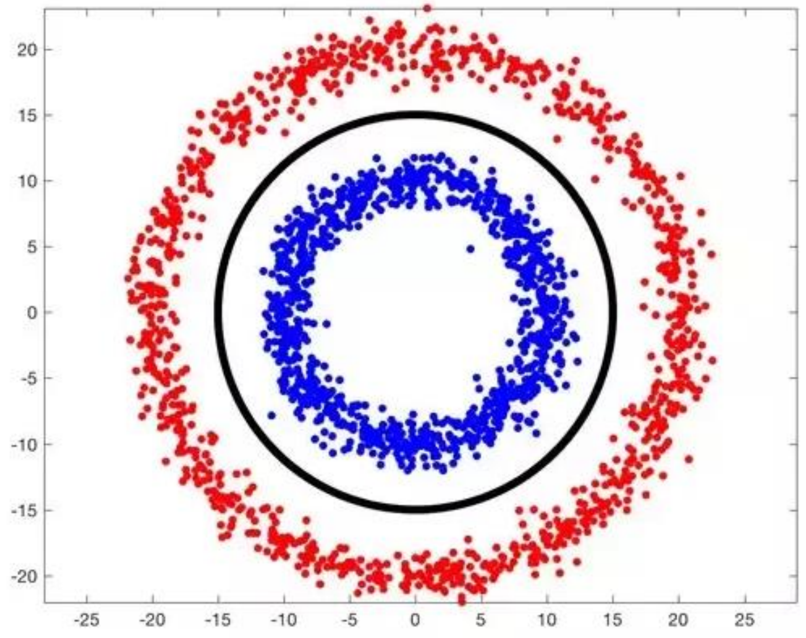
Polynomial kernel, degree=2



RBF Kernel

- Capture complex nonlinear relationships in data.
- Maps input data into an infinite-dimensional feature space using a Gaussian function.

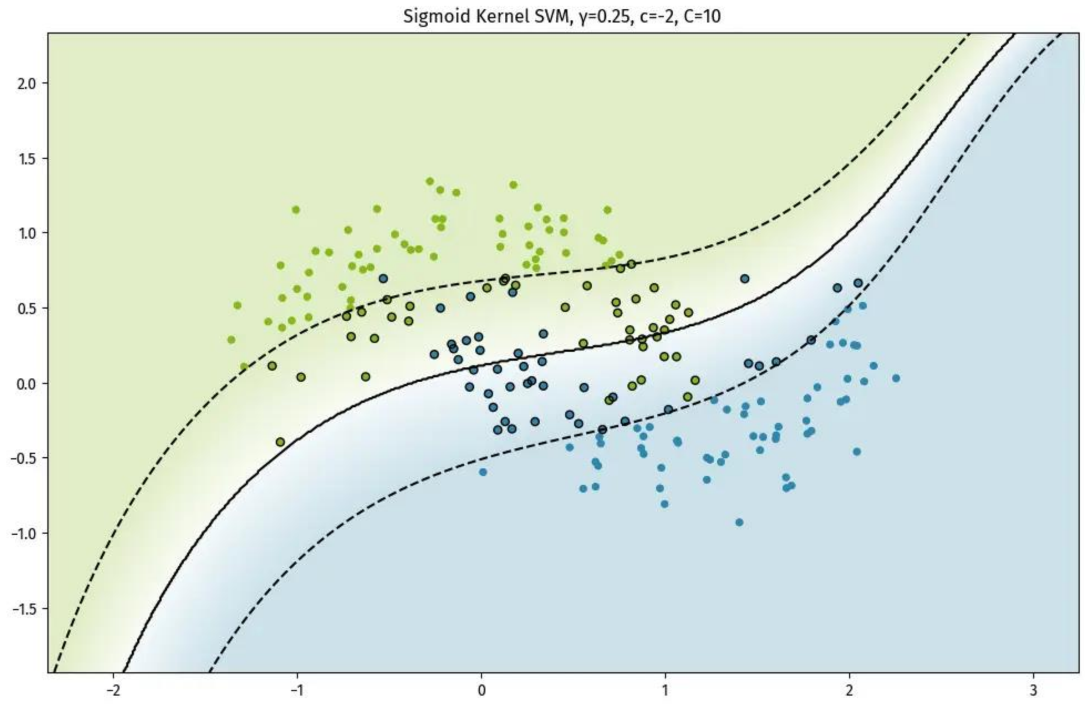
$$k(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(-\gamma \left\| \mathbf{x}_1 - \mathbf{x}_2 \right\|^2\right)$$



Sigmoid Kernel

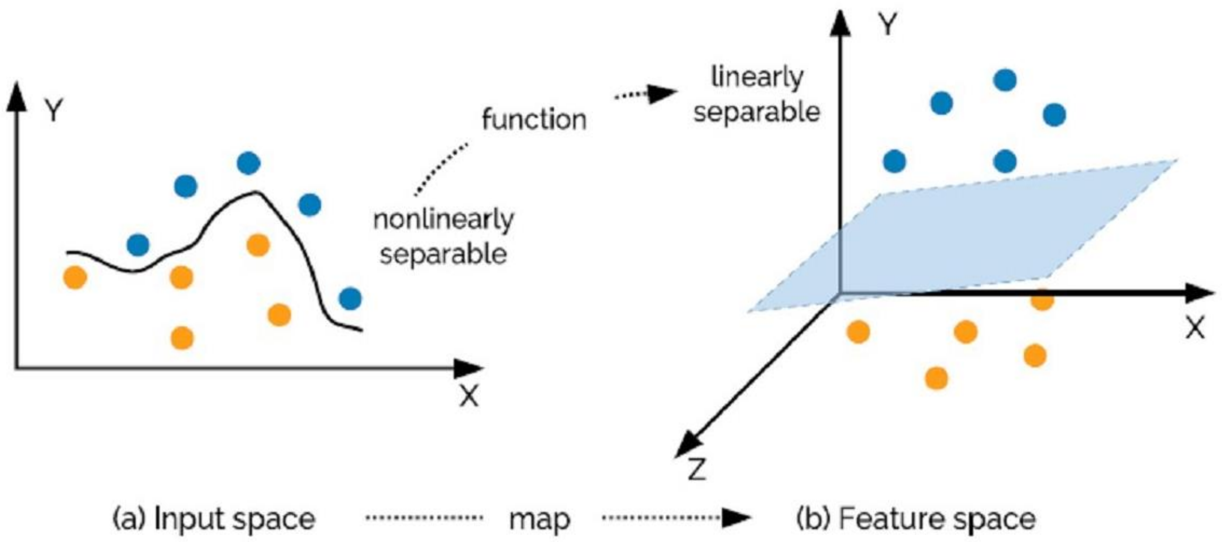
- Similar to the logistic sigmoid function.
- It computes the dot product of two vectors and applies a *tanh* transformation.
- Complex structure and difficult for humans to interpret.

$$k(\mathbf{x}_1, \mathbf{x}_2) = \tanh(\gamma\, \mathbf{x}_1 \cdot \mathbf{x}_2 + c)$$



Kernel Trick

- Transforms low-dimensional input space into a higher-dimensional space.
- Converts non-separable problems into separable problems.
- An [example visualization]() from YouTube.



Advantages of SVM

Works on both linear and non-linear data.

Not sensitive to outliers.

Works really well with a clear margin of separation.

Effective in high-dimensional spaces, where the number of dimensions is greater than the number of samples.

Memory efficient as it uses a subset of the training set (support vectors).

Effective on datasets with multiple features, like financial or medical data.

Different kernel functions can be specified for the decision function.

Possible to specify custom kernels.

Disadvantages of SVM

Training complexity depends on the size of the dataset. Not suitable for large datasets.

Requires a long training time.

Does not perform well when the dataset has more noise.

Produce suboptimal results with imbalanced datasets.

Implementation

- SVM implementation from Scikit-Learn library.
- For classification -
  - SVC
  - NuSVC
  - LinearSVC
- For regression -
  - SVR
  - NuSVR
  - LinearSVR

Summary

- Learned what SVM is and its usage
- Basic terminologies
- Kernel trick
- Pros and cons of SVM
- Code…

Class Notebook

- The notebook was created and executed in kaggle (https://www.kaggle.com/)
- Uses a publicly available dataset, *UniversalBank*

- Dataset Link: https://www.kaggle.com/datasets/nishantchaudhary07/universalbank
- Link to the notebook: https://www.kaggle.com/amrito/class-7-svm-code

Assignment

- Purpose: Learning to use SVM using a regression task.

- Task:
  - Explore and preprocess the dataset
  - Perform feature encoding and scaling if necessary
  - Build your model using SVR from sklearn
  - Measure performance (Try to maximize it!)

- Dataset: *Loan_Data.csv* from class 5 (EDA)
- Target column name: ***LoanAmount***
- Performance metric: r2 or coefficient of determination

Resource

- SVM Tutorials:
  - https://www.freecodecamp.org/news/svm-machine-learning-tutorial-what-is-the-support-vector-machine-algorithm-explained-with-code-examples/
  - https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf

- Kernel Functions:
  - https://data-flair.training/blogs/svm-kernel-functions/
  - https://techvidvan.com/tutorials/svm-kernel-functions/
  - https://pythongeeks.org/svm-kernel-function/

- Hyperparameter Tuning:
  - https://www.geeksforgeeks.org/svm-hyperparameter-tuning-using-gridsearchcv-ml/
  - https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

- K-fold Cross Validation:
  - https://machinelearningmastery.com/k-fold-cross-validation/

Resource

- Courses from Andrew Ng:
  - https://www.youtube.com/watch?v=Rm-XWhI5-Pc&t=1s (old one, recommended for building good intuition)
  - https://www.youtube.com/playlist?list=PLkDaE6sCZn6FNC6YRfRQc_FbeQrF8BwGI

- GeeksForGeeks: https://www.geeksforgeeks.org/machine-learning/

- Sign up for GitHub Student Developer Pack and get tons of free resources, including 3 months of Datacamp access. Follow one of the specializations from Datacamp's Skill Tracks or Career Tracks.

Thank you