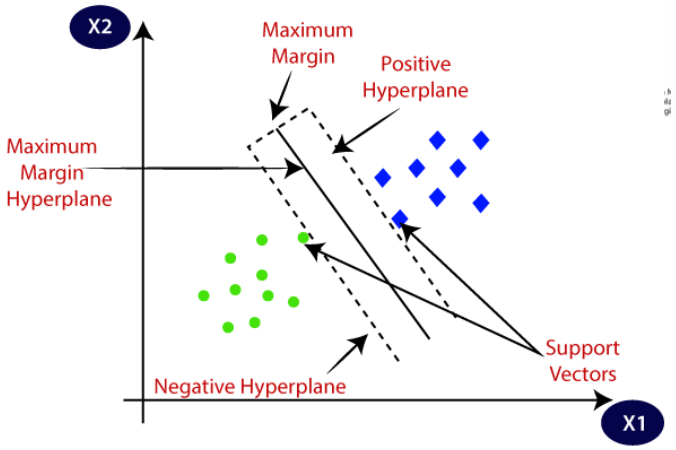
CSE4207 - Pattern Recognition

Zamil-8B-483

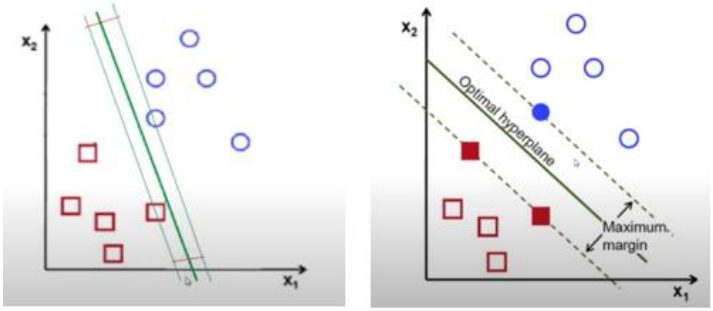
Lecture 03 – Supervised Learning

Support Vector Machines

- Support Vector Machines, a supervised machine learning algorithm used for classification and regression tasks and effective in dealing with complex, high-dimensional datasets.
- SVM performs classification by first transforming the training records into higher dimensions. Then, within the new dimension, it searches for the best decision boundary that separates the training records of one class from others.
- The main objective of SVMs is to find the best hyperplane that separates the data points of different classes in a way that maximizes the margin, or the distance, between the classes. This hyperplane serves as the decision boundary for classifying new, unseen data points.



- Maximum marginal hyperplane is calculated from the records that fall in the hyperplane pair (called "Support Vectors") to be used as the decision boundary.



- From the Figure, we see that the records distributed between two class values are linearly separable; however, there can be an infinite number of lines separating them.
- The target of an SVM is to find the best line (best decision boundary) that will help minimize classification error for unseen records.

Non-linear SVM

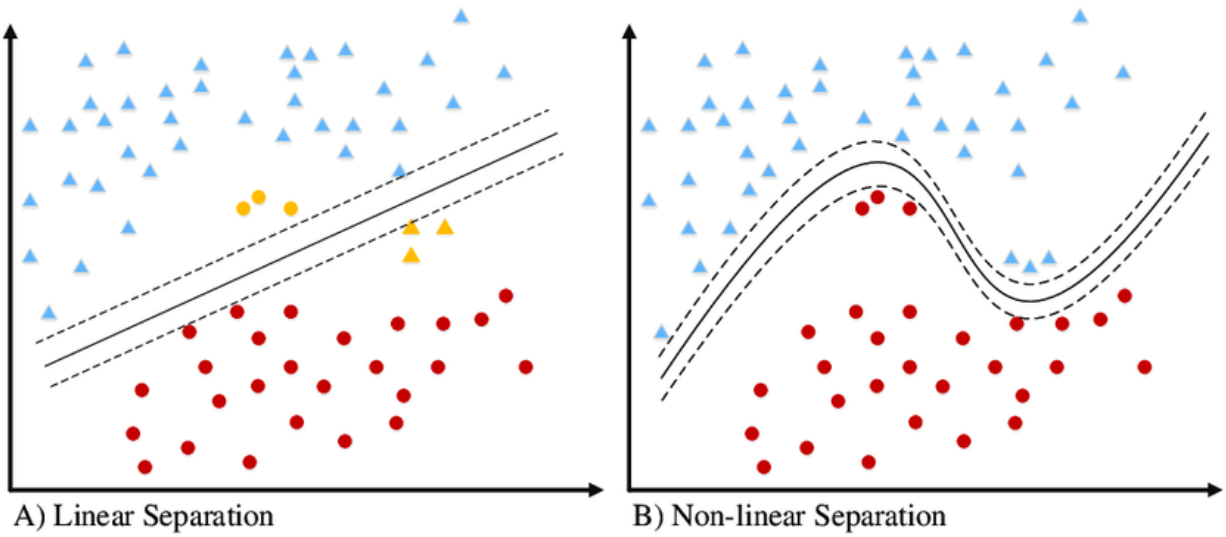
- Non-linear SVMs extend the concept of SVMs to handle non-linear decision boundaries by mapping the input features into a higher-dimensional space.
- This is done through a process called the kernel trick.
- The kernel trick allows the algorithm to compute the dot product of the data points in this higher-dimensional space without explicitly calculating the transformation, saving computational resources.
- The choice of kernel function determines the shape of the decision boundary.

Different types of kernel

1. Linear Kernel
2. Polynomial Kernel
3. Radial Basis Function (RBF) or Gaussian Kernel

Linear vs Non-linear SVM

When we can easily separate data with hyperplane by drawing a straight line is Linear SVM. When we cannot separate data with a straight line we use Non–Linear SVM.



Linear SVM	Non-Linear SVM
It can be easily separated with a linear line.	It cannot be easily separated with a linear line.
Data is classified with the help of hyperplane.	We use Kernels to make non-separable data into separable data.
Data can be easily classified by drawing a straight line.	We map data into high dimensional space to classify.

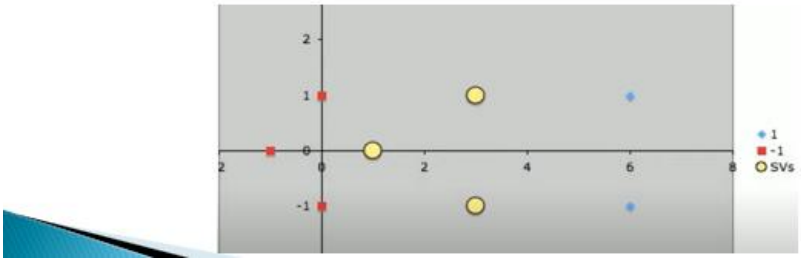
Problem:1(Linear SVM)

Solution:

Here are the data point:

$\binom{3}{1}, \binom{3}{-1}, \binom{6}{1}, \binom{6}{-1}, \binom{1}{0}, \binom{0}{1}, \binom{0}{-1}, \binom{-1}{0}$

Find out the appropriate decision boundary among these data.



By inspection, it should be obvious that there are **three** support vectors,

$$\left\{s_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, s_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix}, s_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}\right\}$$

Each vector is augmented with a 1 as a bias input

So, $s_1 = \binom{1}{0}$, then $\tilde{s_1} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$

Similarly,

$S_2 = \binom{3}{1}, \tilde{S}_2 = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$ $S_3 = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \tilde{S}_3 = \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}$

$$\alpha_1 \tilde{s_1} \cdot \tilde{s_1} + \alpha_2 \tilde{s_2} \cdot \tilde{s_1} + \alpha_3 \tilde{s_3} \cdot \tilde{s_1}$$
$$\alpha_1 \tilde{s_1} \cdot \tilde{s_2} + \alpha_2 \tilde{s_2} \cdot \tilde{s_2} + \alpha_3 \tilde{s_3} \cdot \tilde{s_2}$$
$$\alpha_1 \tilde{s_1} \cdot \tilde{s_3} + \alpha_2 \tilde{s_2} \cdot \tilde{s_3} + \alpha_3 \tilde{s_3} \cdot \tilde{s_3}$$

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = -1$$

$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix} = 1$$

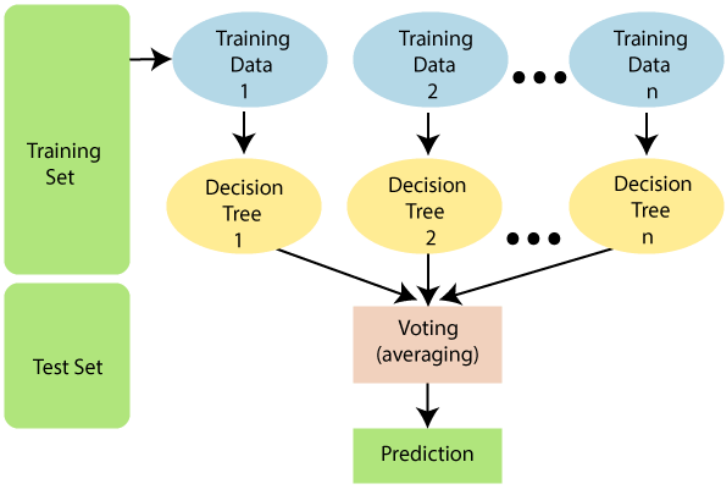
$$\alpha_1 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix} = 1$$

Ensemble learning

- Ensemble learning refers to the technique of combining the predictions of multiple models to improve overall performance and generalization.
- The idea is that by aggregating the predictions of diverse models, the ensemble can often achieve better results than any individual model.
- Ensemble methods are commonly used in various machine learning tasks, including classification, regression, and anomaly detection.
- There are two main types of ensemble learning: **bagging** and **boosting**.
- **Bagging (bootstrap aggregating)**: Creating a different training subset from sample training data with replacement is called Bagging. The final output is based on majority votinag.
- **Boosting**: Combining weak learners into strong learners by creating sequential models such that the final model has the highest accuracy is called Boosting. Example: ADA BOOST, XG BOOST.

Random Forest

- Random Forest is an ensemble learning algorithm that combines the predictions of multiple decision trees to improve the overall accuracy and can handle complex problems.
- It contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.
- Rather than depending on one tree it takes the prediction from each tree and based on the majority votes of predictions, predicts the final output. For regression, it's the average of the predictions.
- It uses bagging methods.



Regression

- Regression determines the statistical relationship between a dependent variable and one or more independent variables which are used to predict real or continuous values.
- The ultimate goal of the regression algorithm is to plot a best-fit line or a curve between the data.
- There are several types of regression. Linear Regression, Multiple Linear Regression, and Polynomial Regression.

- In simple words, "Regression shows a line or curve that passes through all the data points on target-predictor graph in such a way that the vertical distance between the data points and the regression line is minimum."
- The distance between data points and line tells whether a model has captured a strong relationship or not.
- **Example:** Suppose there is a marketing company A, who does various advertisement every year and get sales on that. The below list shows the advertisement made by the company in the last 5 years and the corresponding sales:

Advertisement	Sales
$90	$1000
$120	$1300
$150	$1800
$100	$1200
$130	$1380
$200	??

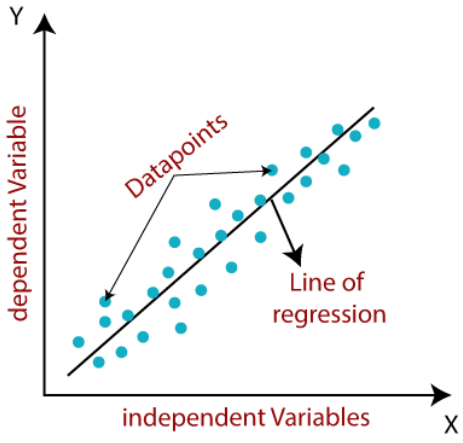
- Now, the company wants to do an advertisement for $200 in the year 2019 and wants to know the prediction about the sales for this year. So to solve such type of prediction problems in machine learning, we need regression analysis.

Application of regression

- Forecasting continuous outcomes like house prices, stock prices, or sales.
- Predicting the success of future retail sales or marketing campaigns to ensure resources are used effectively.
- Predicting customer or user trends, such as on streaming services or e-commerce websites.
- Analyzing datasets to establish the relationships between variables and output.
- Predicting interest rates or stock prices from a variety of factors.
- Creating time series visualizations.

Simple Linear Regression

- Linear regression finds the linear relationship between the dependent variable and one independent variable using a best-fit straight line.
- Generally, a linear model predicts by simply computing a weighted sum of the input features, plus a constant called the bias term (also called the intercept term).
- In this technique, the dependent variable is continuous, the independent variable(s) can be continuous or discrete, and the nature of the regression line is linear.



Multiple Linear regression

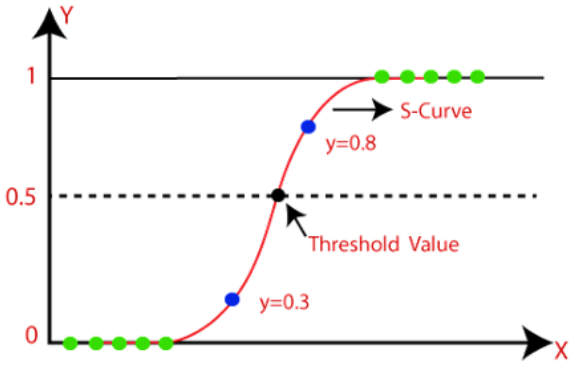
- Multiple linear regression refers to a technique that uses two or more independent variables to predict the outcome of a dependent variable.
- It achieves a better fit in the comparison to simple linear regression when multiple independent variables are involved.
- The equation for multiple linear regression is similar to the equation for a simple linear equation, i.e., $y(x) = p0 + p1x1$ plus the additional weights and inputs for the different features which are represented by b_nx_n.
- The formula for multiple linear regression would look like, $y(x) = b_0 + b_1x_1 + b_2x_2 + \ldots + b_nx_n$

Math

Logistic Regression

- Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance belongs to a given class or not.
- It is a statistical method that is used for building machine learning models where the dependent variable is dichotomous: i.e. binary which works with categorical variables such as 0 or 1, Yes or No, True or False, Spam or not spam, etc.
- It has the ability to provide probabilities and classify new data using continuous and discrete datasets.
- Logistic regression uses the **sigmoid function** which is a mathematical function used to map the predicted values to probabilities.
- It maps any real value into another value within a range of 0 and 1 and forms a curve like the "S" form that curve is called the Sigmoid function or the logistic function.
- Probability is either 0 or 1, depending on whether the event happens or not.
- For binary predictions, you can divide the population into two groups with a cut-off of 0.5. If the hypothesis is above 0.5 is considered to belong to group A, and below is considered to belong to Group B.

sigmoid function: $p(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$



Type of Logistic Regression

On the basis of the categories, Logistic Regression can be classified into three types:

1. **Binomial:** In binomial Logistic regression, there can be only two possible types of dependent variables, such as 0 or 1, Pass or Fail, etc.
2. **Multinomial:** In multinomial Logistic regression, there can be 3 or more possible unordered types of the dependent variable, such as “cat”, “dogs”, or “sheep”. In this case, the softmax function is used in place of the sigmoid function.
3. **Ordinal:** In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as “low”, “Medium”, or “High”.

Application of LR

- In health care, logistic regression can be used to predict if a tumor is likely to be benign or malignant.
- In the financial industry, logistic regression can be used to predict if a transaction is fraudulent or not.
- In marketing, logistic regression can be used to predict if a targeted audience will respond or not.