

# Support Vector Machine

Support Vector Machines (SVM) are supervised ML algorithm that are used for classification and regression. SVM are typically used for classification, regression and outlier detection purpose. An SVM builds a model that assigns new data point to one of the given categories. SVM can be used for linear classification purpose.

In [ ]:

1

## SVM terminology

In [ ]:

1

### Hyperplane

A hyperplane is a decision boundary which separates given set of data points having different class labels. The SVM classifier separates the data point using a hyperplane with maximum amount of margin.

This hyperplane is known as the maximum margin hyperplane and the linear classifier it defines is known as the maximum margin classifier.

In [ ]:

1

### Margin

A margin is a separation gap between two lines on the closest data point.

It is calculated as the perpendicular distance from the line to closest data point.

In SVM we try to maximize this separation gap so that we get maximum margin

In [ ]:

1

In SVM our main objective is to select a hyperplane with the maximum possible margin between SVM in the given dataset. SVM searches for the maximum margin hyperplane in the following 2 step process:

1) Generate hyperplane which segregates the classes in the best possible way. There are many hyperplane that might classify the data. We should look for the best hyperplane that represent the largest separation or margin between two classes.

2) We choose the hyperplane so that the distance from it to the support vector on each side is maximized. If such a hyperplane exists it is known as maximum margin hyperplane and its classifier is known as maximum margin classifier

In [ ]:

1

In [ ]:

1

### Problems with Dispersed data

Quite Often the data points are so dispersed that it is not possible to separate them using a linear hyperplane

In such situation SVM uses a kernel trick to transform the the input space to a higher dimensional space, It uses a mapping function to transform the 2-D input space into 3-D input space.

We can easily segregate the data points using linear separation.

In [ ]:

1

### Kernel Trick

SVM algorithm is implemented using a kernel, It uses a technique called kernel trick.

In simple words a kernel is just a function that maps the data to a higher dimension where data is separable.

A kernel transforms a low dimensional input data space into a higher dimensional space. So it converts non linear separable problem into a linear separable problem by adding more dimensions into it.

hence the kernel trick helps us to build a more accurate classifier. It is useful in separation problems

In [ ]:

1

### Different Type of Kernel Tricks

- 1) Linear Kernel
- 2) Polynomial Kernel
- 3) Radial Basis function (RBF) -- Gaussian Kernel
- 4) Sigmoid Kernel

In [ ]:

1

### Linear Kernel

**linear kernel formula:  $K(x_i, x_j) = x_i^T x_j$**

Linear kernel is used when the data is linearly separated. It means the data can be separated using a single line. It is one of the most common used type of kernel tricks, mostly used when there are large number of features in a data set.

Linear kernel is often used for text classification purpose.

Training with linear kernel is usually faster because we need to optimize the regularization parameters.

1

### Polynomial Kernel Tricks

Polynomial kernel represents the similarity of vectors (training sample) in a feature space over polynomials of the original variables.

The polynomial kernel looks not only at the given feature of input samples to determine their similarity but also combination of input samples.

It is typically used in NLP (Natural Language Processing).

In [ ]:

1

**Radial Basis function (RBF) -- Gaussian Kernel**

RBF is a general purpose kernel, It is used when we have no prior knowledge about the data.

In [ ]:

1

**Sigmoid Kernel**

Sigmoid kernel has its origin to neural networks, we can use it as a proxy for neural networks.

Sigmoid kernel is given by the following formula

$$K(x,y) = \tanh(axTy+c)$$

In [ ]:

1