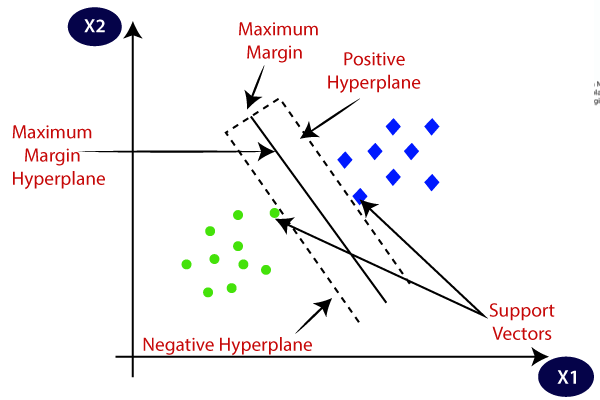
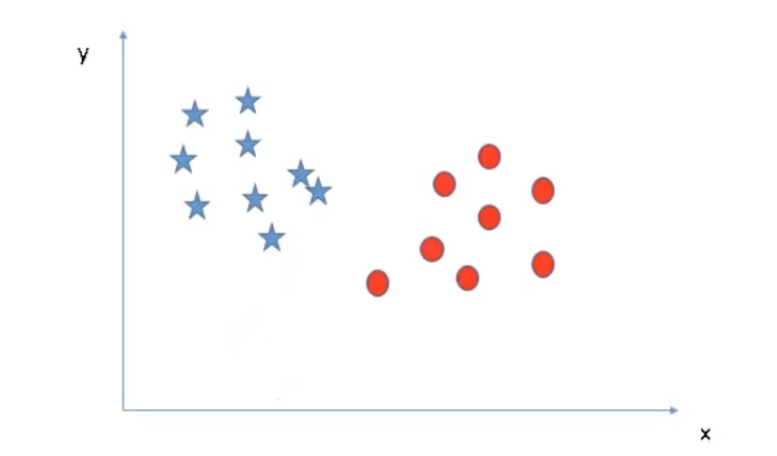
**Support vector machine**

Support Vector Machines (SVM) is a type of supervised machine learning algorithm that can be used for **classification or regression analysis**. SVM works by finding a **hyperplane** (a line or a higher dimensional plane) that separates the data into classes or predicts the target value. The hyperplane is selected such that it **maximizes the margin** between the closest data points (**support vectors**) of the two classes. These support vectors are the most important samples and have the greatest impact on the position of the hyperplane. The SVM algorithm uses these support vectors to make predictions for new data. **SVM algorithm can be used for Face detection, image classification, text categorization, etc**.

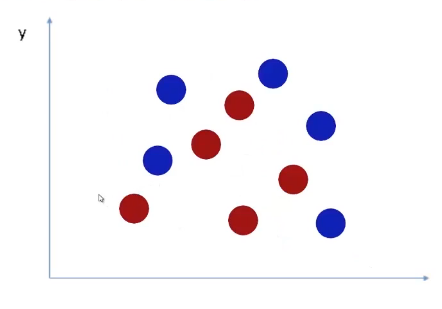


**SVM can be of two types:**

**Linear SVM:** Linear SVM is used for **linearly separable data**, which means if a dataset can be classified into two classes by using a **single straight line**, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier. **Ex:**



**Non-linear SVM:** Non-Linear SVM is used for **non-linearly separated data**, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier. **Ex:**



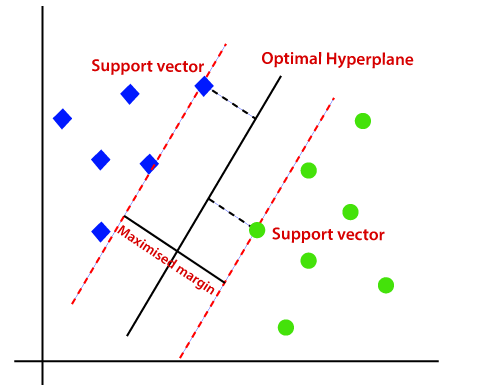
**Some important terms in the SVM algorithm:**

**Hyperplane:** There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the **best decision boundary** that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

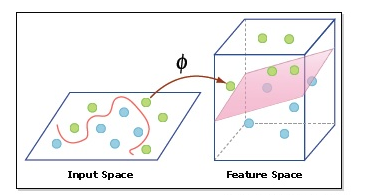
The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane. We always create a hyperplane that has a **maximum margin**, which means the **maximum distance between the data points**.

**Support Vectors:** The data points or vectors that are the **closest to the hyperplane** and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

**Margin:** The distance between the vectors and the hyperplane is called as margin. And the goal of SVM is to **maximize this margin**. **The hyperplane with maximum margin is called the optimal hyperplane.**



**Kernel:** A kernel helps us **find a hyperplane in the higher dimensional space** without increasing the computational cost. Usually, the computational cost will increase if the dimension of the data increases. This increase in dimension is required when we are unable to find a separating hyperplane in a given dimension and are required to move in a higher dimension. SVM kernel make low dimension data into higher dimension so that we can draw hyperplane to separate the class.



**Advantage and disadvantage of SVM:**

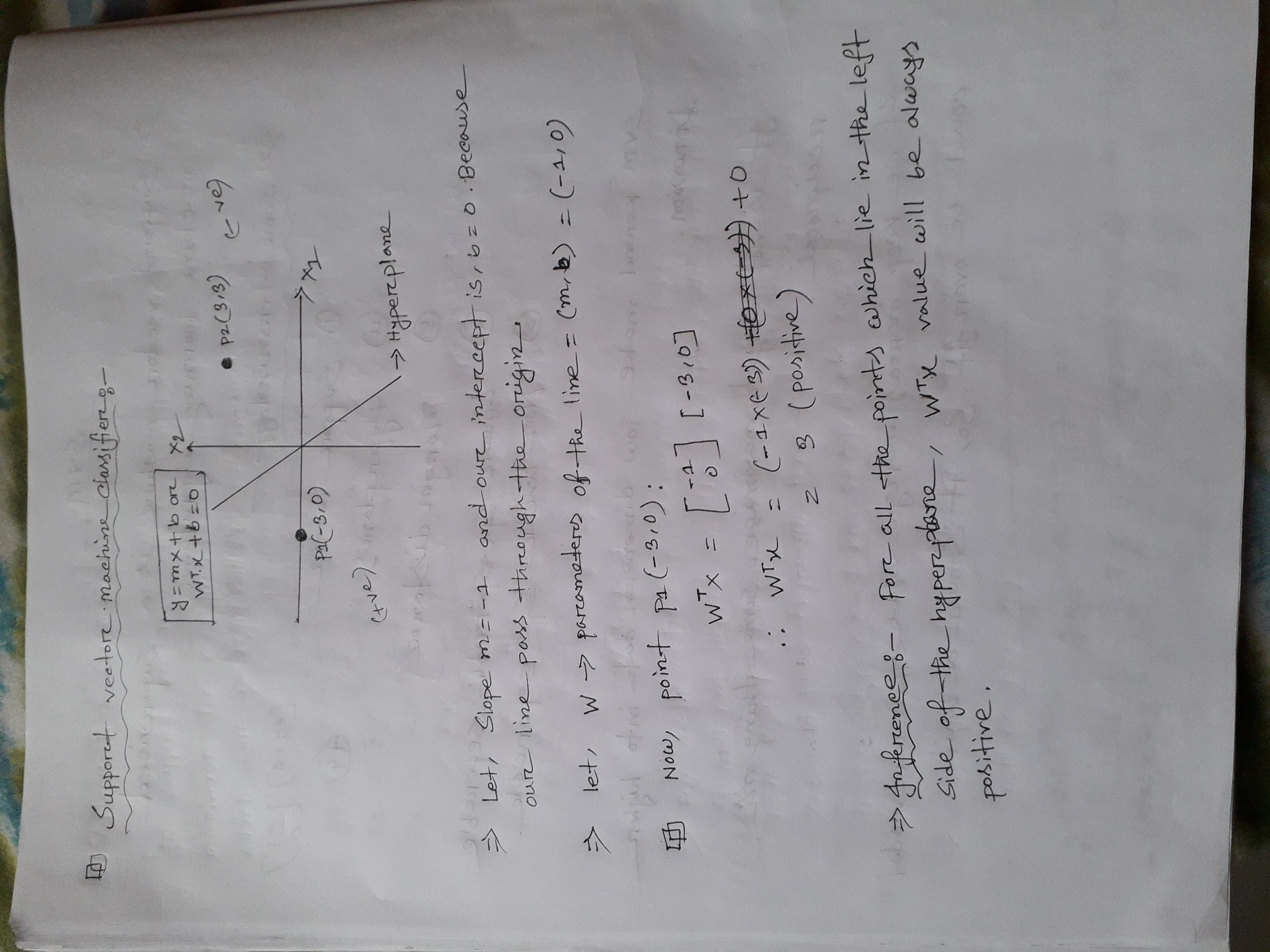
**Advantages:**

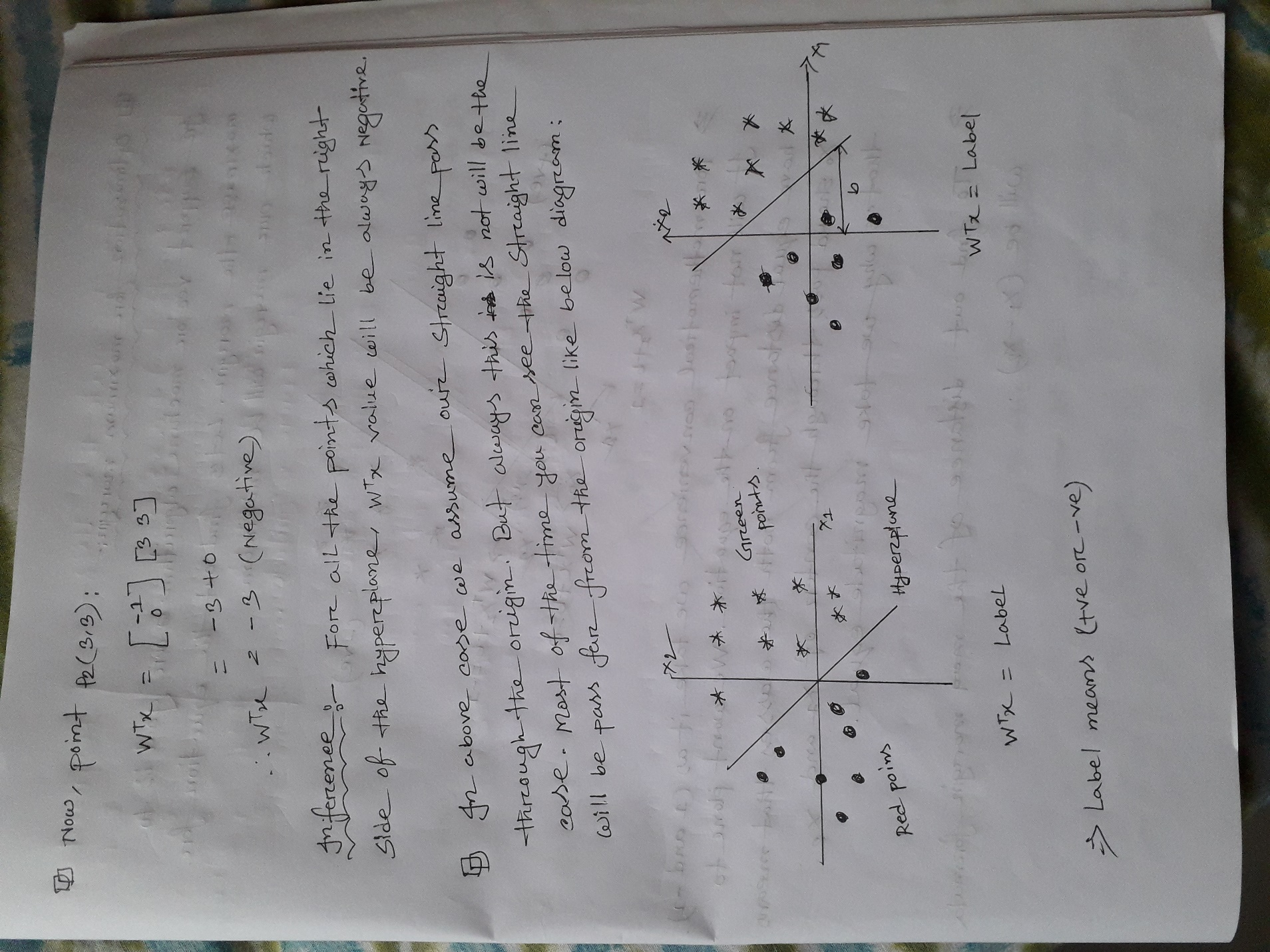
* Works well with smaller datasets.
* Works efficiently when there is a clear margin of separation.
* Works well with high dimensional data.

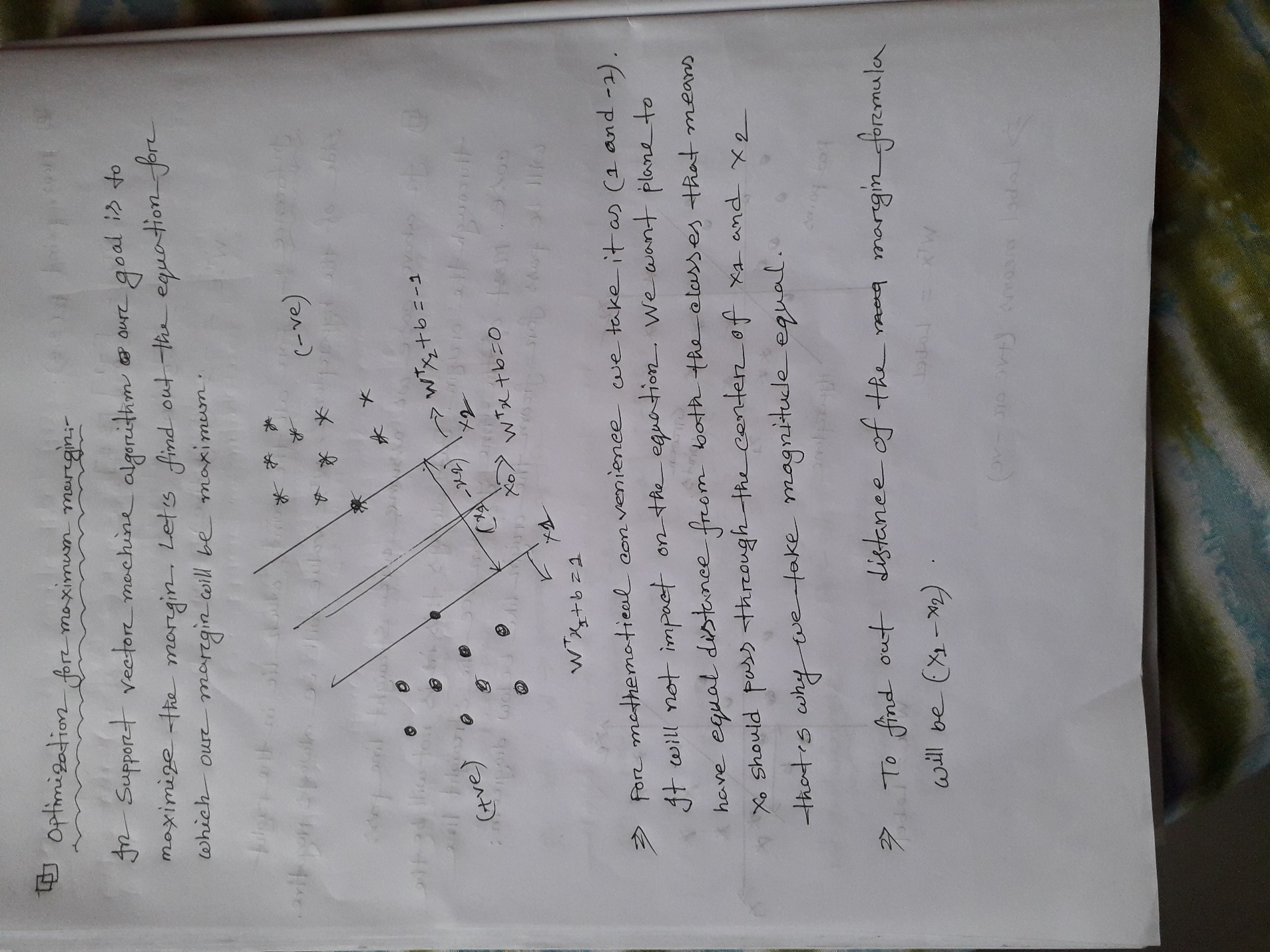
**Disadvantages:**

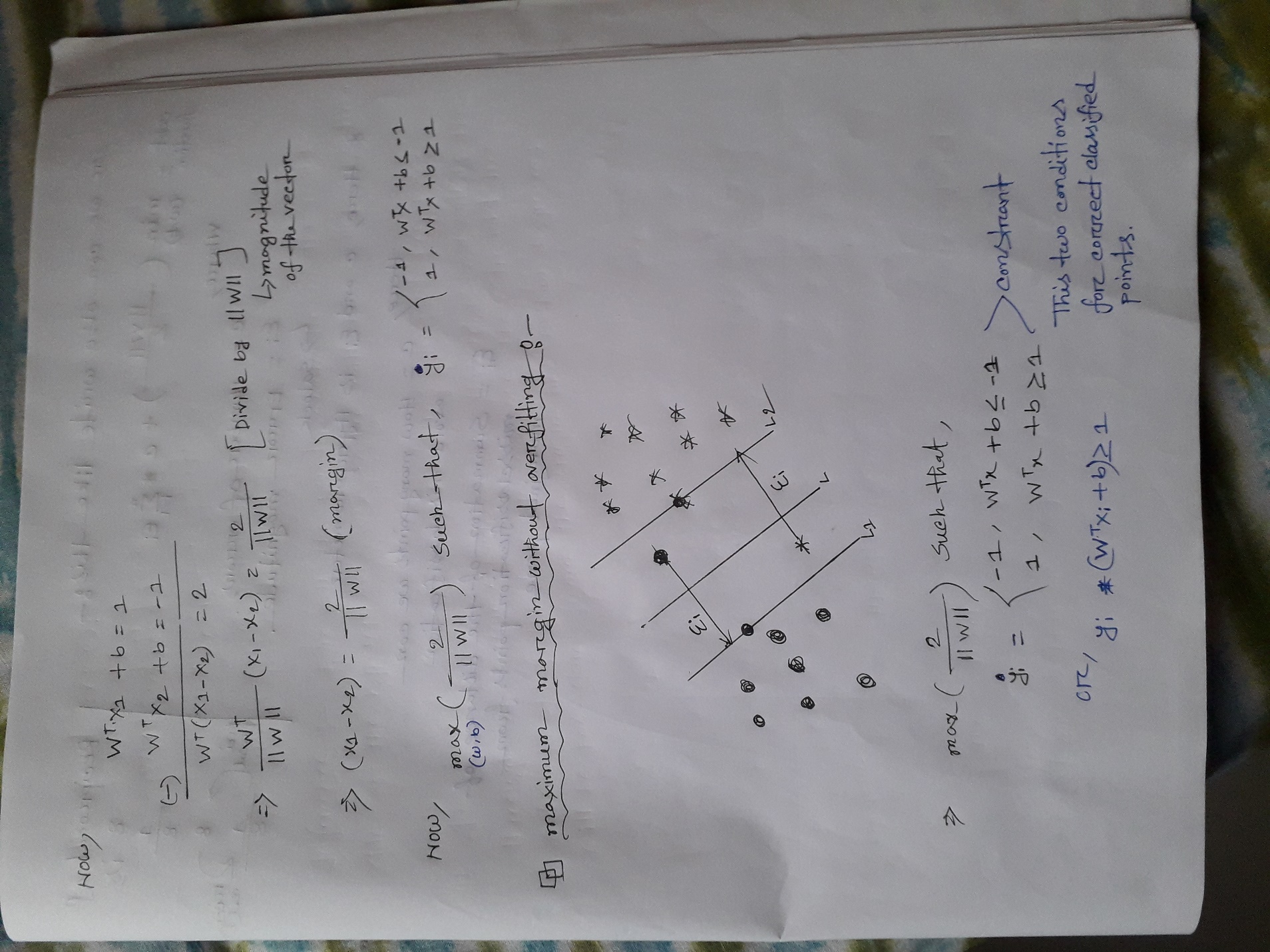
* Not suitable for large datasets as the training time is higher.
* Not suitable for noisier datasets with overlapping classes.

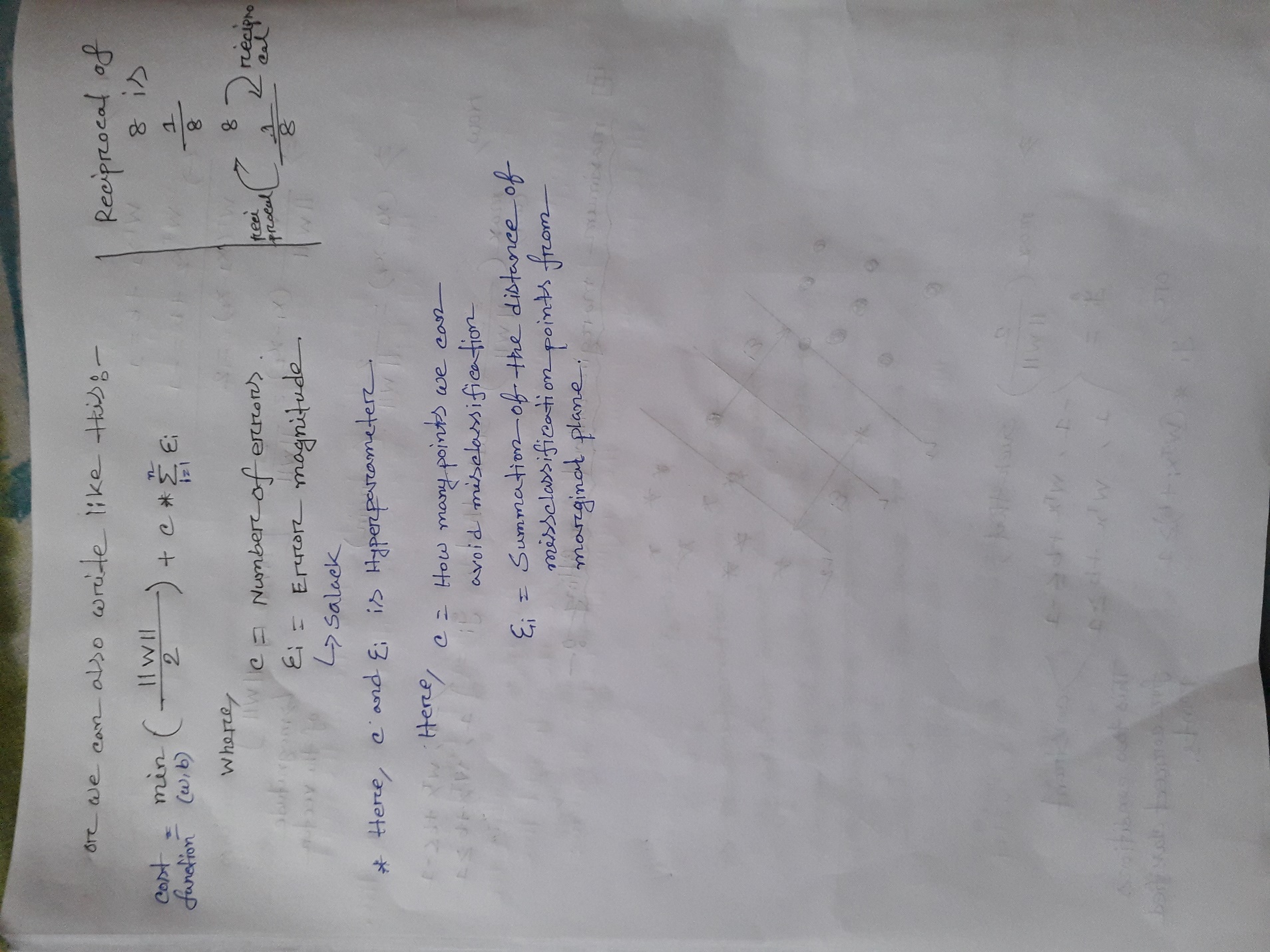
**Support vector machine classifier:**

****

****

****

****

****

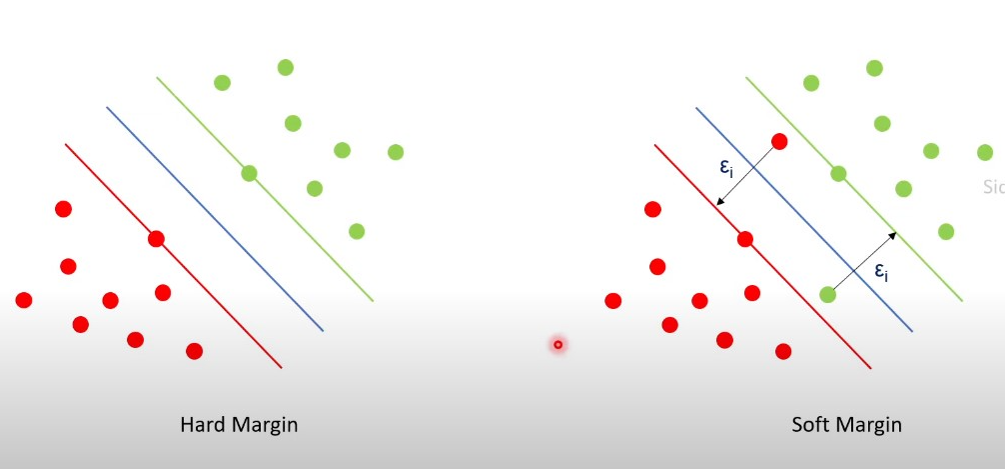
**Hard Margin vs Soft Margin:**

**Hard Margin:**

* When the data is linearly separable, and don’t want to have any misclassification, then use SVM with a hard margin.
* In hard margin we don’t allow any misclassification.

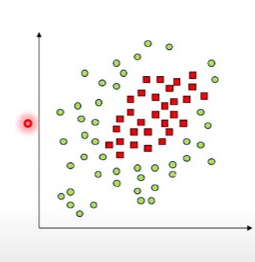
**Soft Margin:**

* When a linear boundary is not feasible or the data is not completely linearly separable or want to allow some misclassification in the hope of achieving better generality, use SVM with soft margin for classifier.
* What soft margin does is it tolerates a few dots to get misclassified and it tries to balance the trade-off between finding a line that maximizes the margin and minimizes the misclassification.



**Non-Linear Support Vector Machine Classifier:**

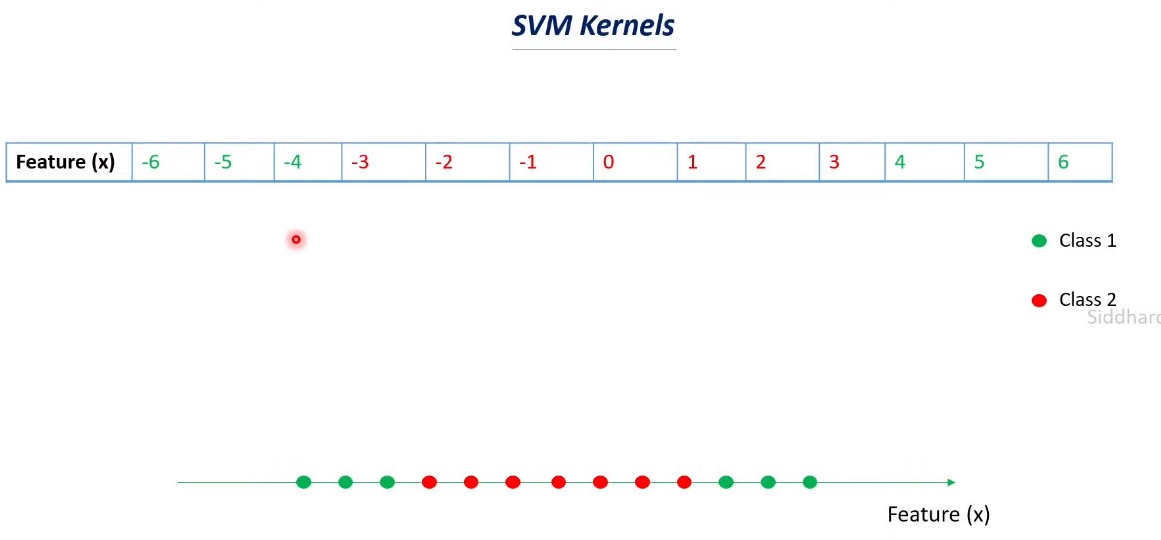
Non-linear support vector machine classifier means the data is not linearly separable. We can’t separate the class just draw a straight line.



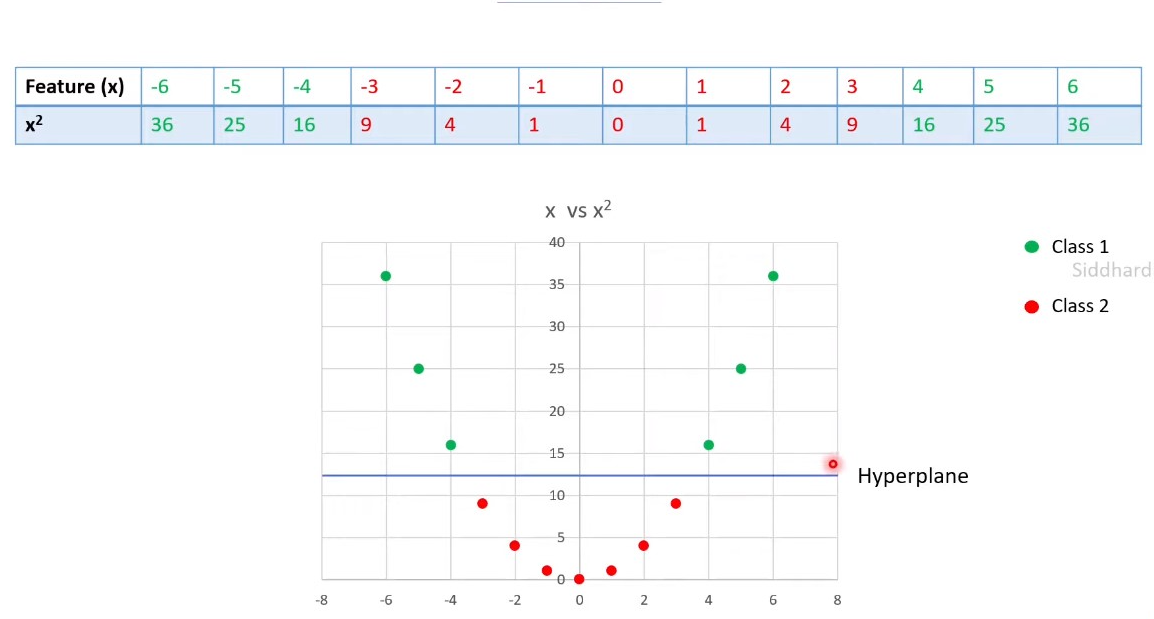
**Types of SVM kernels:**

* Linear kernel
* Polynomial kernel
* Radial Basis function (RBF)
* Sigmoid kernel

**How does kernels work non-linear data let’s see:**



Feature(x) now 1D data. If we plot it then all the points line in a single line and we cannot separate class by draw a straight line right! So, we will increase the dimension of the feature.



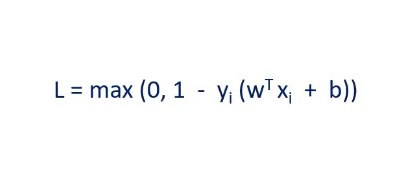
As you can see after increasing the dimension the red and green point will be separable by a hyperplane and this is done the kernels. It’s help to take lower dimension data into higher dimension so that we can separate the classes.

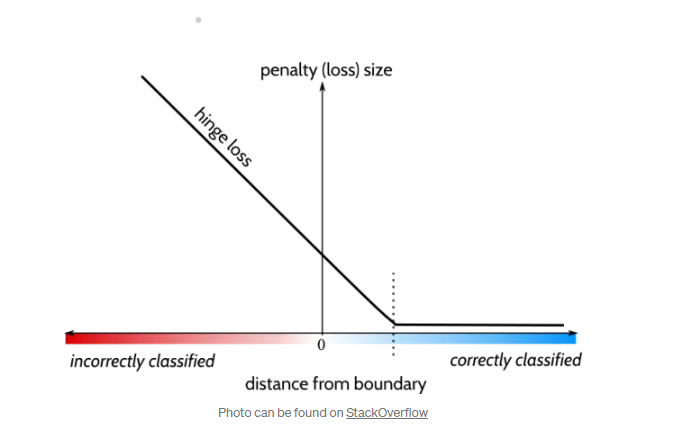
**Formula of different kernels:**

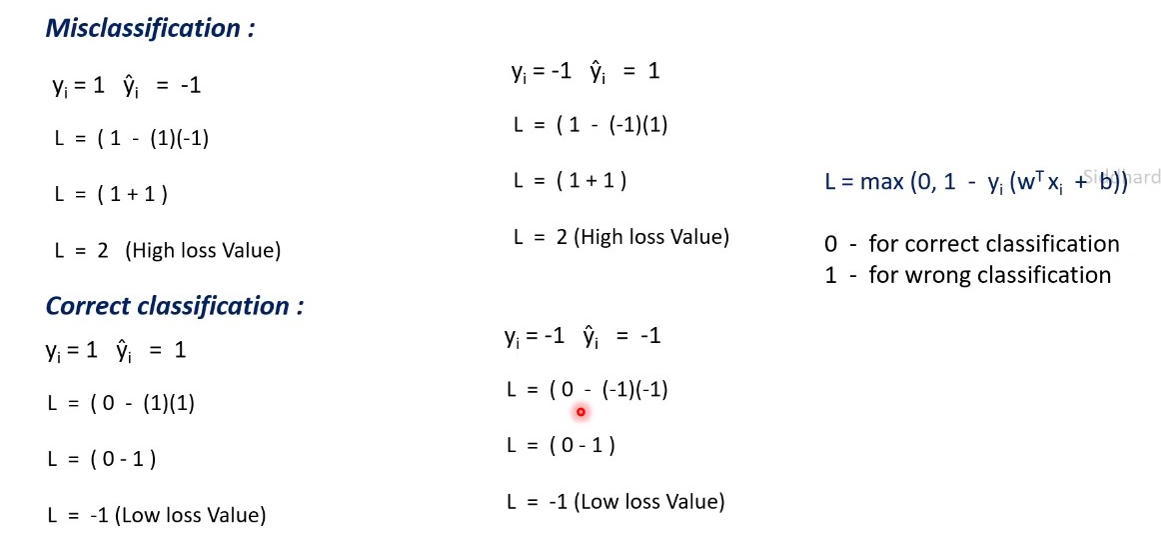
* **Linear kernel**
* **Polynomial kernel**
* **Radial basis function or rbf kernel**
* **Sigmoid kernel**

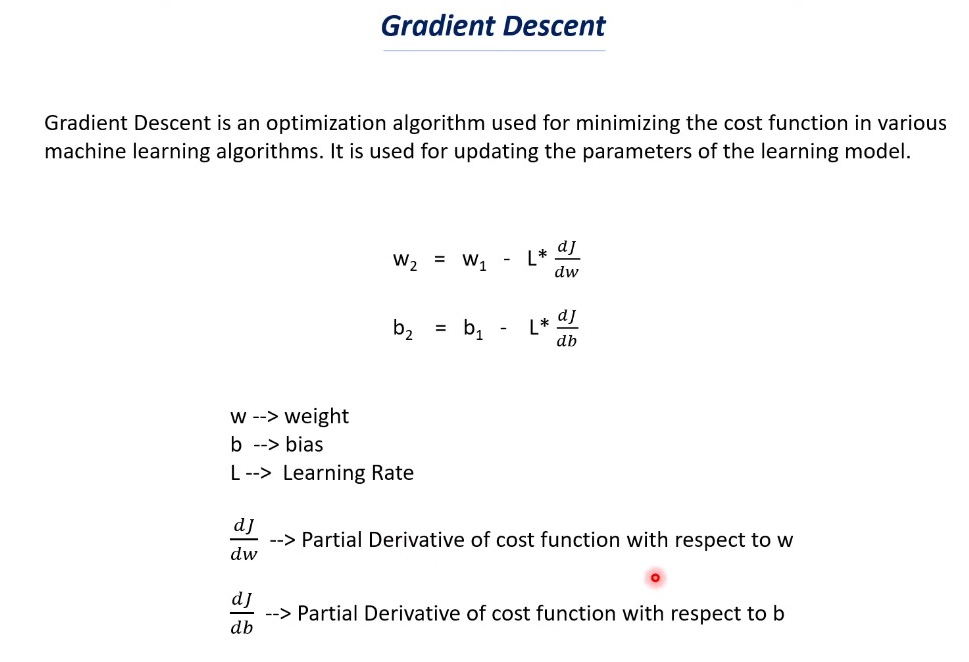
**Loss Function:**

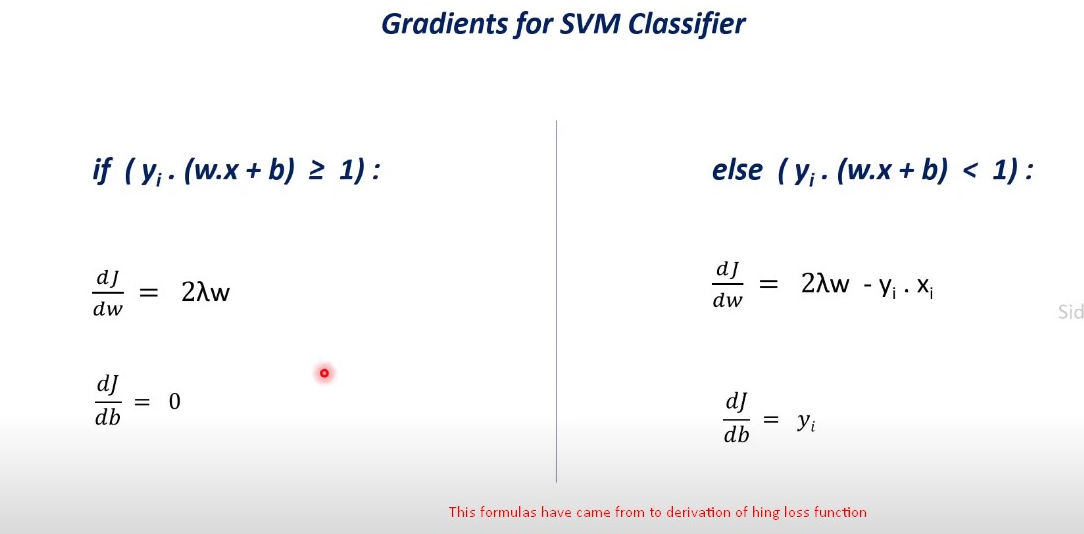
Loss function measures how far an estimated value is from its true value. It is helpful to determine which model performs better and which parameters are better. For support vector machine classifier is used “**Hinge Loss**”as the loss function. **Hinge Loss** is one of the types of loss function, mainly used for maximum margin classification models. Hinge loss incorporates a margin or distance from the classification boundary into the loss calculation**. Even if new observations are classified correctly, they can incur a penalty if the margin from the decision boundary is not large enough**. The hinge loss increases linearly. The hinge loss is mostly associated with soft-margin support vector machines. Formula will be:

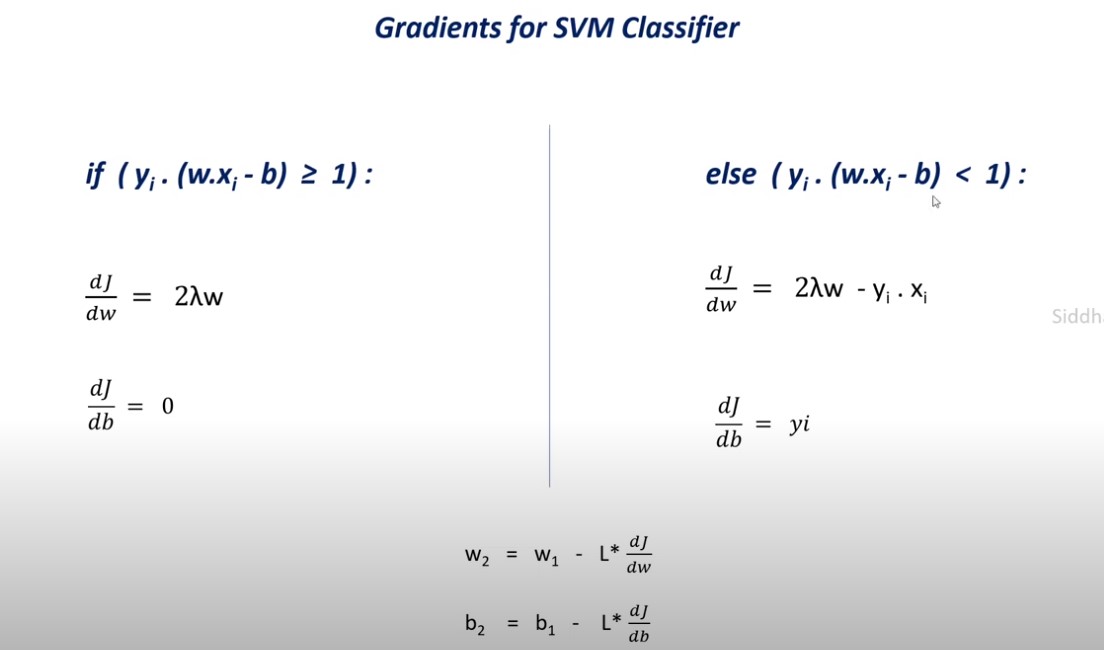












Here, **lambda** is for regularization term.

