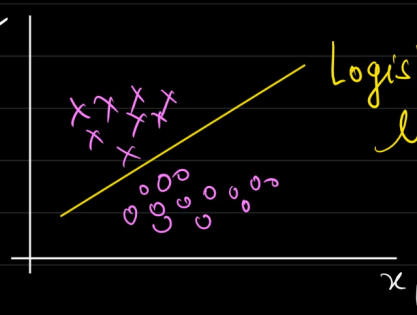
Support Vector machines

① Support Vector classifier ⟶ Classification
② Support Vector Regressor → Regressor

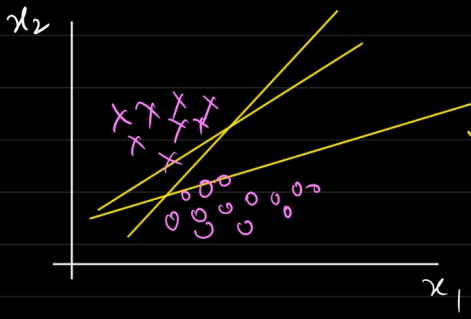
SVC, SVR

① Support Vector classifier

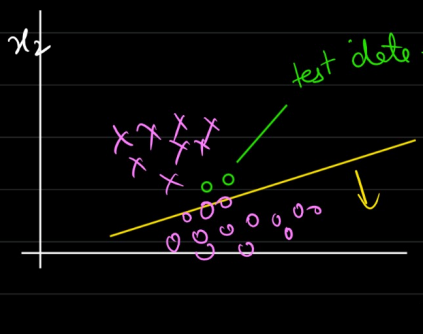
In Log Reg



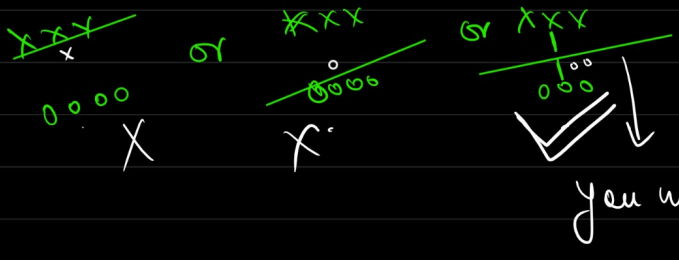
Logistic reg model finds linear decision boundary which separates two classes.



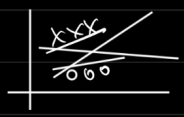
→ Any line can be decision boundary in Logistic Reg.



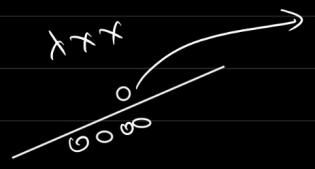
→ Logistic regression decision boundary is prone to error.



* Logistic regression model doesn't care about margin / space across two class.



* if the data points changes slightly, Log reg model will give error.

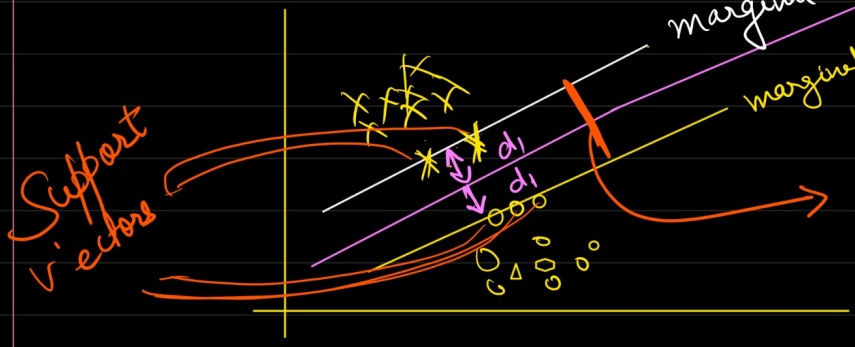


It will belong to other class.

Hence Log reg is prone to error.

* Support Vector classifier

→ Make a decision boundary such that it is equidistant from two or more than two data points of opposite class.



→ best fit line decision boundary

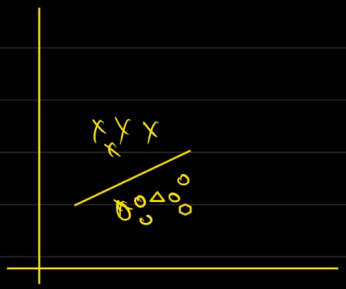
margin → the gap between the nearest datapoints of both class.

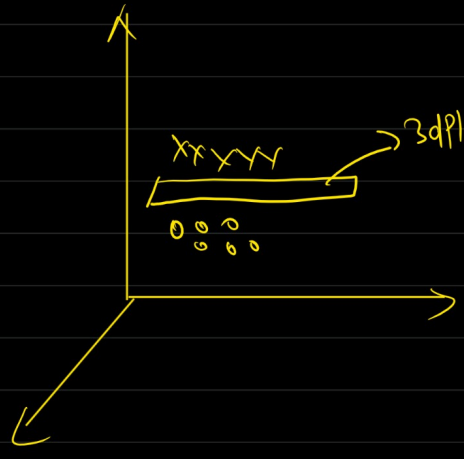
(nearest dp's can be multiple)

⇓

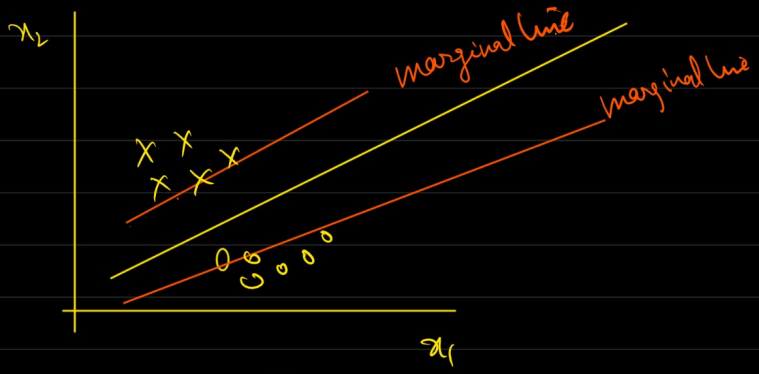
Support Vectors.

Log regression

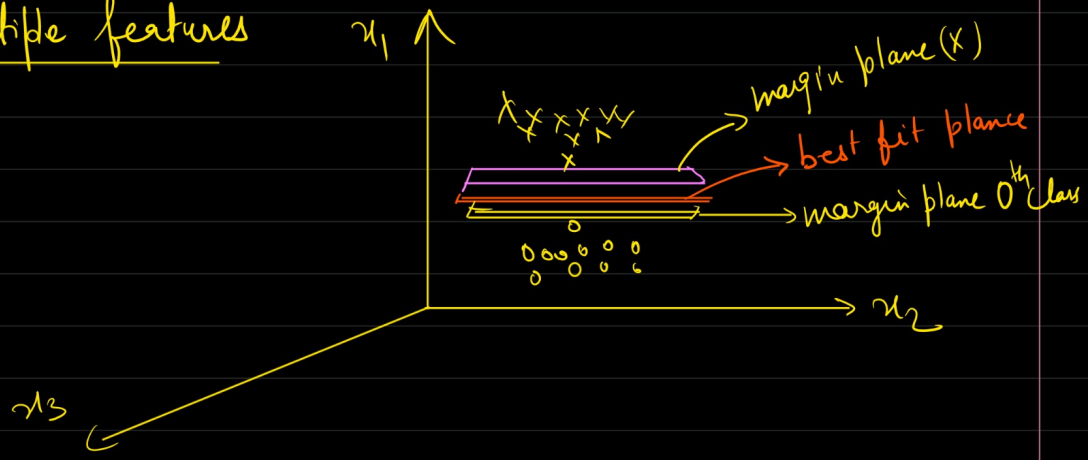


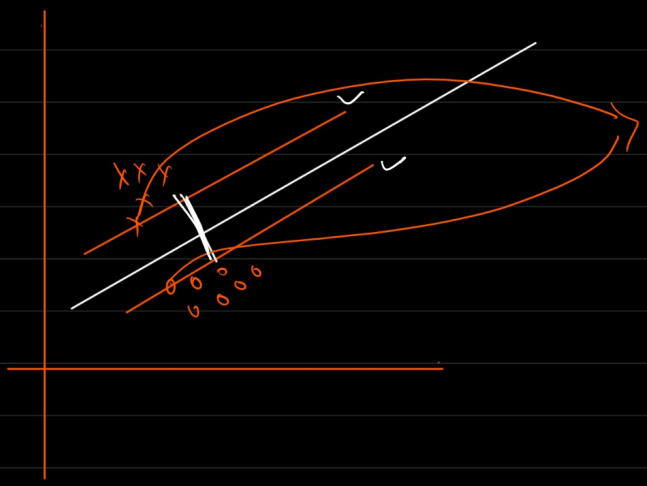


Similarly in SVM



In case of multiple features





Support vectors

↓

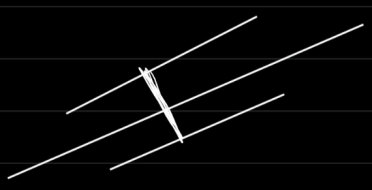
nearest datapoint of either class.

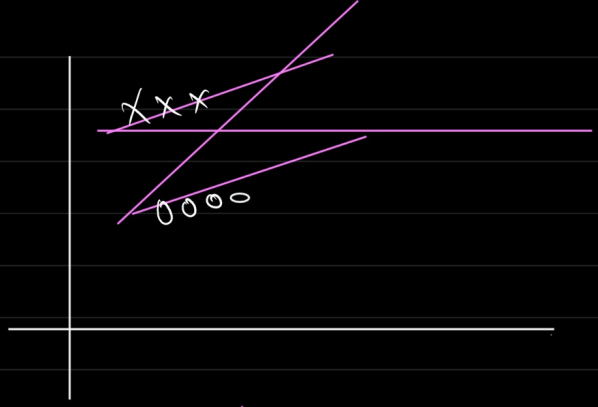
It is called support vectors because these nearest data points helps to create the right classifier

⇒ There are no limitations on support vectors.

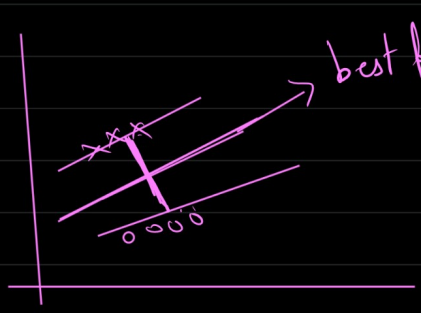
* Min no of support vector (1+1) = 2

* Support vectors will always choose central lines passing exactly through the centre; that's why Support Vector machine is also called margin classifier.

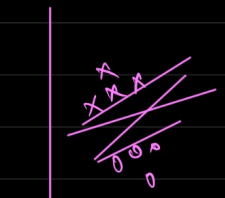




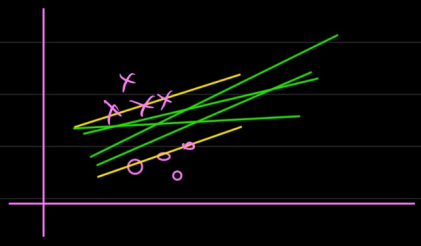
Always

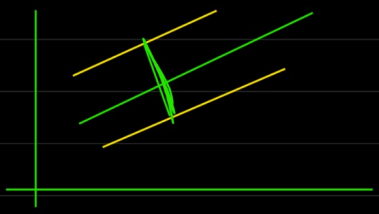


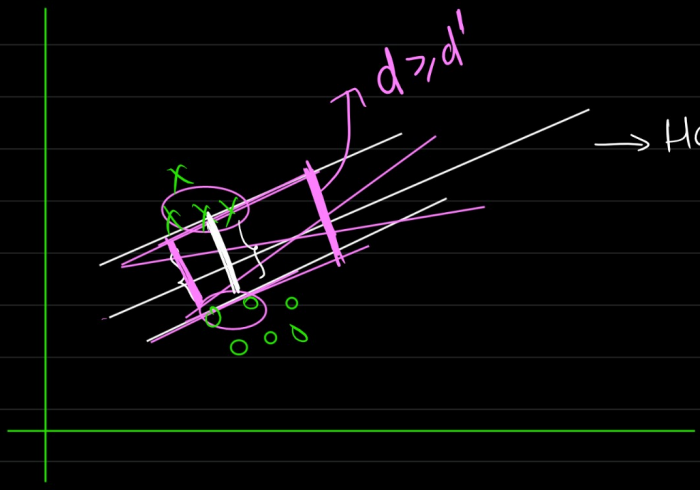
best fit line.



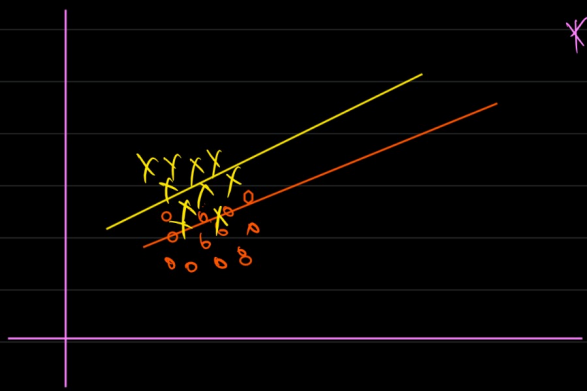
Step-1 - find out all possible classifier
Step-2 Find out margin of each classifier.
Step-3 Select which has maximum margin as classifier



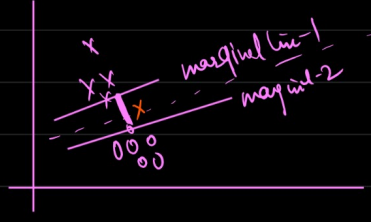




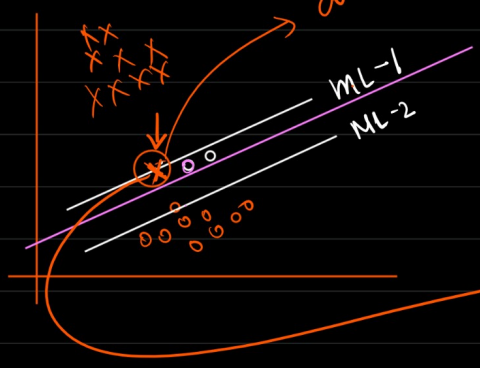
→ Hard margin ⇒ none of the data points are overlapping / misclassified



* Soft margin
⇓
Some data points are misclassified (Error)

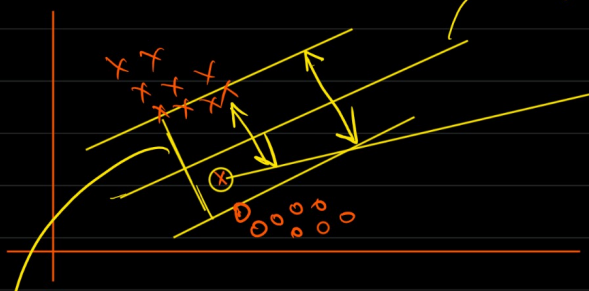


Scen-1



due to this one X, ML-1 is very close to ML-2, that's why best fit line is also very close to class 0.

→ This dp is causing overfitting.



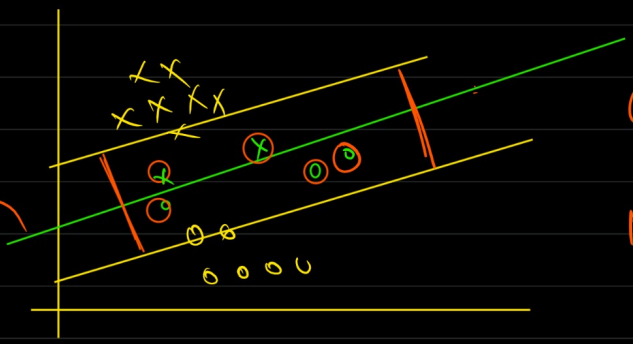
→ misclassified | error. during training.

but the best fit line is more generic.

→ Soft margin = Some points are misclassified

→ you are ready to misclassify some of the data points in order to have good testing accuracy, its a soft margin classifier.

This parameter is 'C' (regularisation parameter)



C = 5

↓

No of datapoints, you sacrificed so that best fit line is not overfitting.

→ linear Support Vector classifier.