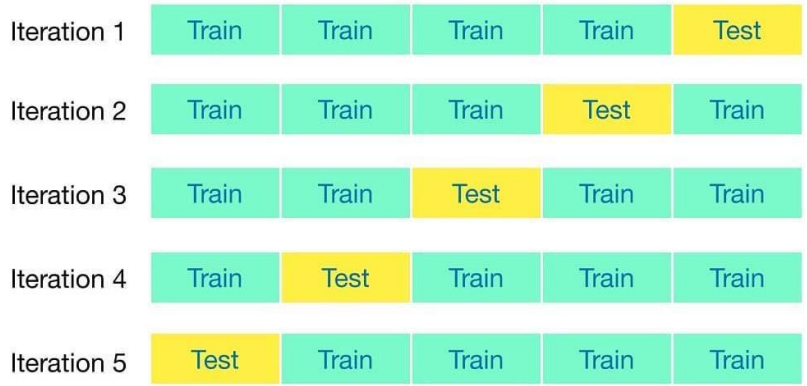
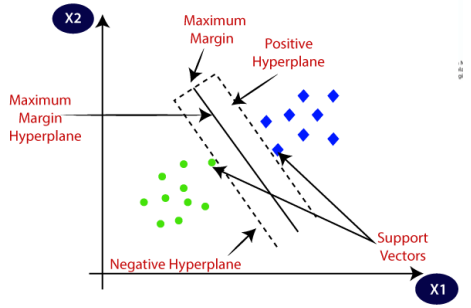
Siddhardhan

K-Fold Cross-Validation

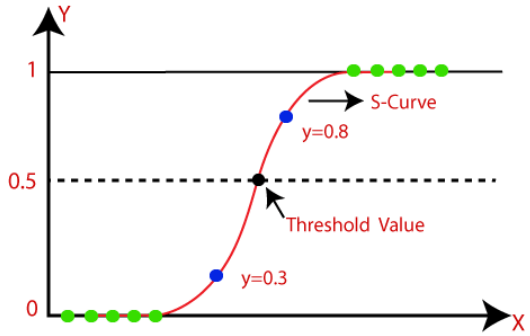


K-Fold Cross-Validation

In K-Fold Cross Validation, we split the dataset into **"K"** number of **folds** (subsets). One chunk of data is used as test data for evaluation & the remaining part of the data is used for training the model. Each time, a different chunk will be used as the test data.

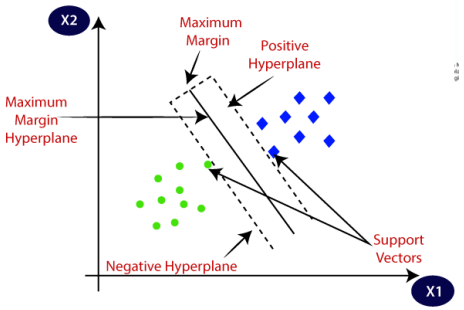


Support Vector Machine



Logistic Regression

K-Fold Cross-Validation



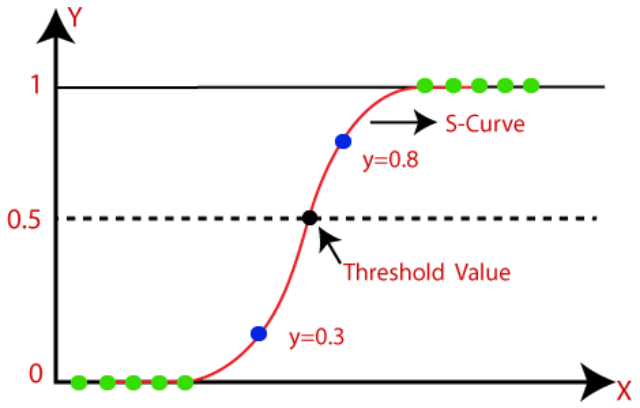
Support Vector Machine

K = 5

	Dataset					*Accuracy*
Iteration 1	Train	Train	Train	Train	Test	88%
Iteration 2	Train	Train	Train	Test	Train	83%
Iteration 3	Train	Train	Test	Train	Train	86%
Iteration 4	Train	Test	Train	Train	Train	81%
Iteration 5	Test	Train	Train	Train	Train	84%

$$\text{Mean Accuracy} = \frac{88 + 83 + 86 + 81 + 84}{5} = 84.4\ \%$$

K-Fold Cross-Validation



Logistic Regression

K = 5

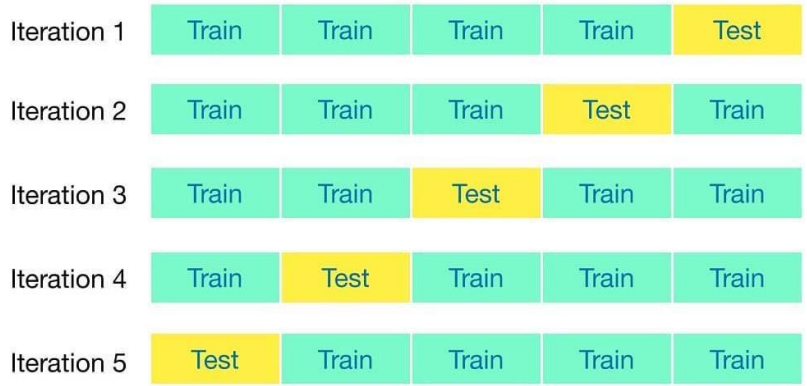
	Dataset					*Accuracy*
Iteration 1	Train	Train	Train	Train	Test	90%
Iteration 2	Train	Train	Train	Test	Train	88%
Iteration 3	Train	Train	Test	Train	Train	86%
Iteration 4	Train	Test	Train	Train	Train	91%
Iteration 5	Test	Train	Train	Train	Train	85%

$$\text{Mean Accuracy} = \frac{90 + 88 + 86 + 91 + 85}{5} = 88\ \%$$

K-Fold Cross-Validation

✓ *Accuracy score for SVM = 84.4 %*

✓ *Accuracy score for Logistic Regression = 88 %*



Advantages of using K-Fold Cross-validation:

- Better alternative for train-test split when the dataset is small
- Better for multiclass classification problems
- More reliable
- Useful for Model Selection