So, for a large dataset size, or large size of parameters it is wise to use RandomizedSizeCV.

Logistic Regression For Multiple Class Classification:

In multiclass classification there can be more than 2 values in the target feature.

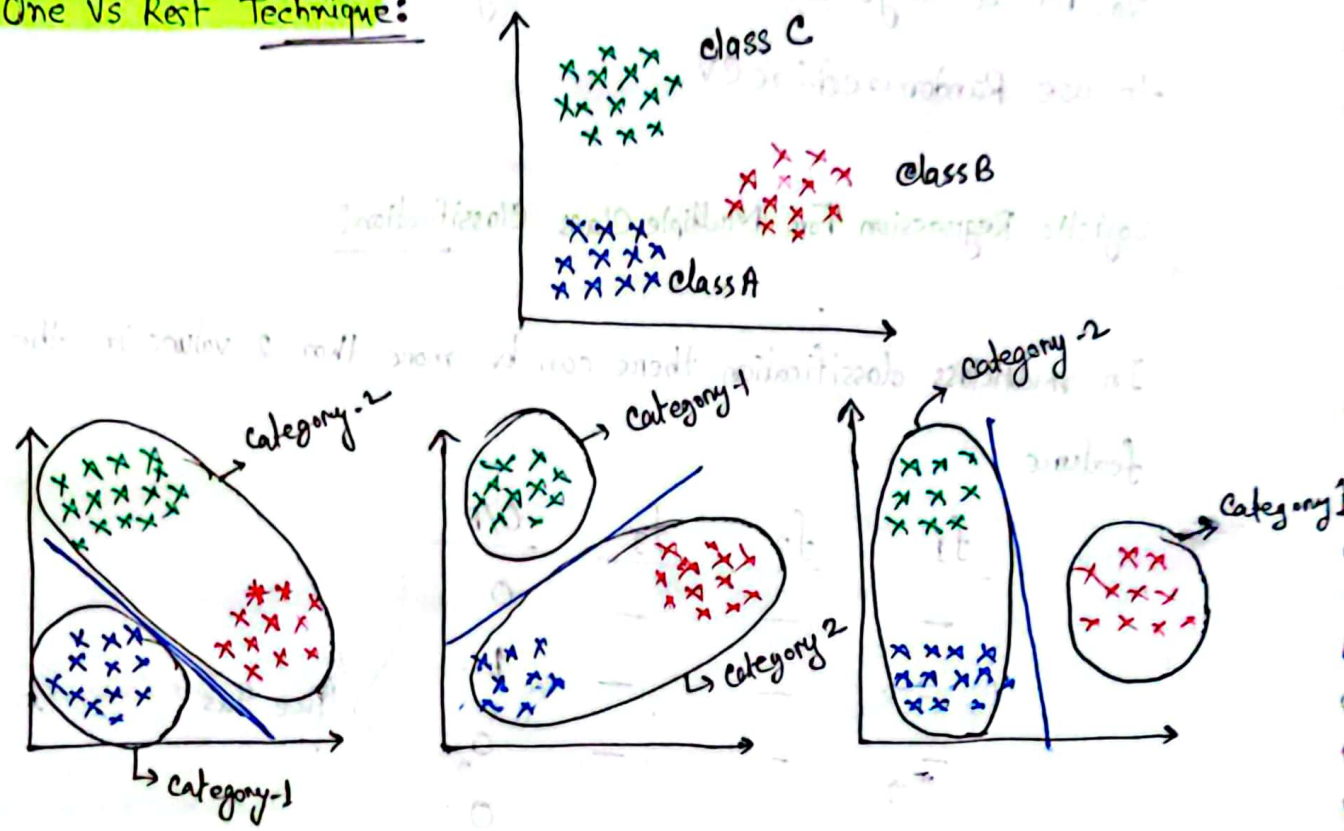
f_1	f_2	f_3	O/P
–	–	–	0
–	–	–	1
–	–	–	0
–	–	–	0
–	–	–	2
–	–	–	1

Like this

We use two types of techniques to solve Multiplass Classification:

1. One versus Rest (OVR)
2. Multinomial

One Vs Rest Technique:



These category will convert the multiclass classification to Binary classification means as here there are 3 classes, this technique will create 3 models. In the first Model (M1), class A will be 1 category, and class B and C combinedly will make another category then model training and prediction will be so done. Linke this in second Model (M2), class B C will be 1 category and class A, B will combinedly be a category. Similarly Model 3 (M3) will do the similar task.

M_1 → Binary Classification

M_2 → Binary classification

M_3 → Binary Classification

~~Then will choose the model which will provide best prediction accuracy.~~

Then we will do one hot encoding (n new features for n independent features)

f_1	f_2	f_3	O/p	O_1	O_2	O_3
-	-	-	0	0	1	0
-	-	-	1	0	0	1
-	-	-	2	1	0	1
-	-	-	2	1	1	0
-	-	-	1	0	0	1
-	-	-	0	1	0	0

for, M1 Model → Input f_1, f_2, f_3 → O/p O_1

M_2 Model → Input f_1, f_2, f_3 → O/p O_2

M_3 Model → Input f_1, f_2, f_3 → O/p O_3

For, new Datapoints M_1, M_2 and M_3 will output some probabilities

M_1 M_2 M_3

[0.20 0.35 0.45] → These are called Multinomial probability

which probability will be max, that model's output will be counted

like here O_3 values will considered as prediction.

While defining model for multiclass classification, we have to use →

LogisticRegression(multi_class = "multinomial", solver = 'lbfgs')

or, LogisticRegression(multi_class = 'multinomial', solver = 'lbfgs')