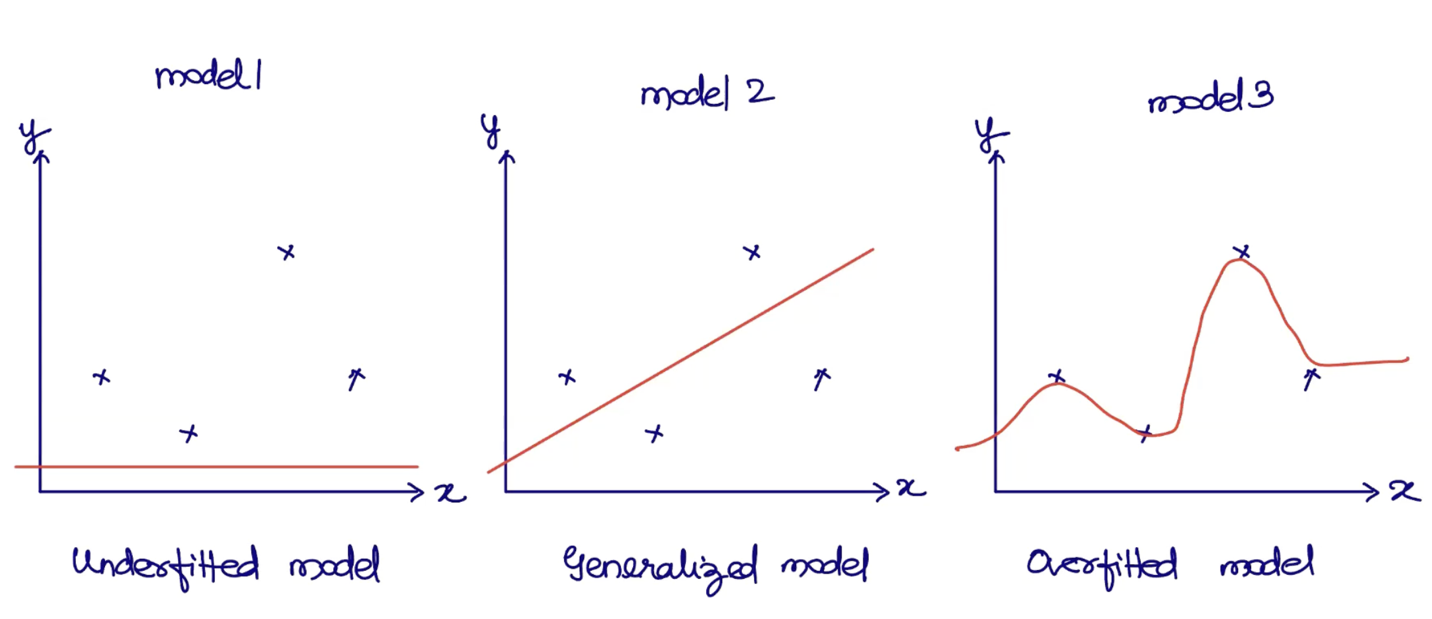
**Ensemble learning**

**Bagging**

When we build a model, it might happen one of these situation.



What is the solution for underfitted model or overfitted model?

One way is, create a for loop and play with random state and use if condition and try to get the best model and the result is a generalized model.

There are other ways that you can achieve a generalized model.

In underfitted model: the error is more and the accuracy is less.

The difference between actual and predicted value is the error.

BIAS VARIANCE TRADEOFF:

What if I get overfitted models everytime? Is there a way to fine-tune the model such that remove the overfitting? Such that I tell the model “don’t memorize, instead understand the pattern”? 🡺 This is where the **BIAS VARIANCE TRADEOFF** comes to the picture.

**BIAS VARIANCE TRADEOFF is** understanding the trade-off between the ability of the model to converge training data whereas the ability of the model to generalize new and unseen data.

**BIAS** here means ability of the model to converge training data.

**VARIANCE** here means making model behavior generalized to deal with unseen data.

You cannot have a model with heavy bias or heavy variance. You somewhere need bagging ( balancing). How to balance this behavior practically?

High variance is like overfitted model. 🡺 **if train Score > test Score , test Score > CL**

Low variance is like underfitted model. 🡺 if **test Score < CL**

Usually the underfitted model call as a HIGH BIAS Model.

Usually the overfitted model call as a HIGH Variance Model.

What HIGH Variance means?

It means the model works BEST with known data. It may not work good with unknown data.

What HIGH BIAS means?

It means creating simplified model which captures some part of training data.

The problem with a simplified model is it doesn’t consider entire training data.

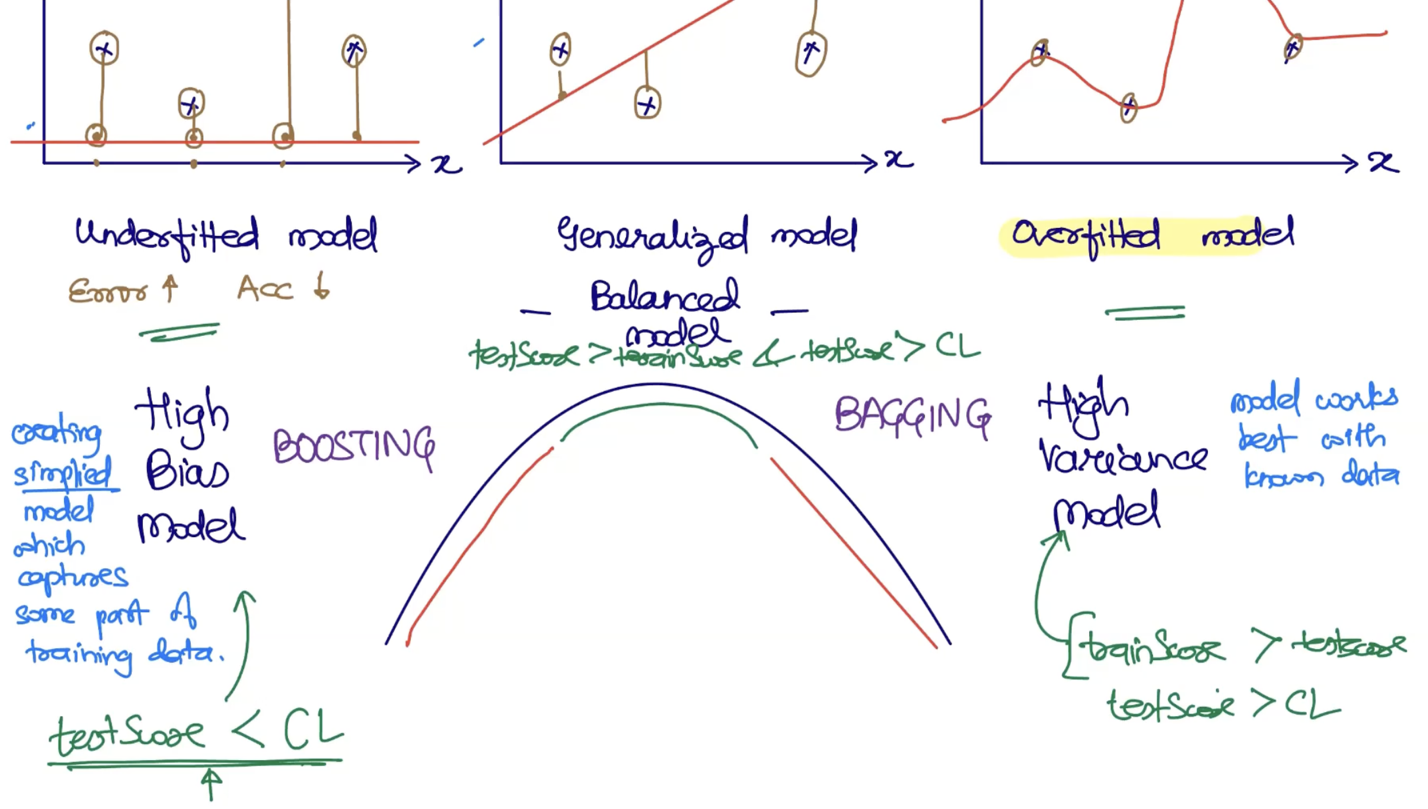
Our goal is, to find a Balance model. A little bit of both (bias and variance)

To fix overfitting model 🡺 use **Bagging algorithms 🡺** force the model not to memorize

To fix underfitting model 🡺 use **Boosting algorithms 🡺** force the model use more sample data to converge

When use bagging? Convert overfitted model to a balanced model

When use boosting? Convert underfitted model to a balanced model



**Ensemble Learning:**

It is all about using a set of weak learners/models for prediction.

There are 3 methods of **Ensemble Learning:**

* Bagging method
* Boosting method
* Stacking method

**Bagging** 🡺 (Bootstrap aggregation)

* Applicable for dealing with over-fitted model

Example:

In Classification [ Email Spam Classification]

Imagine we have a Dataset 🡺 100 records. Label 🡺 spam (50 records) | ham(50 records)

In Bagging algorithm:

There are two questions need to be answered:

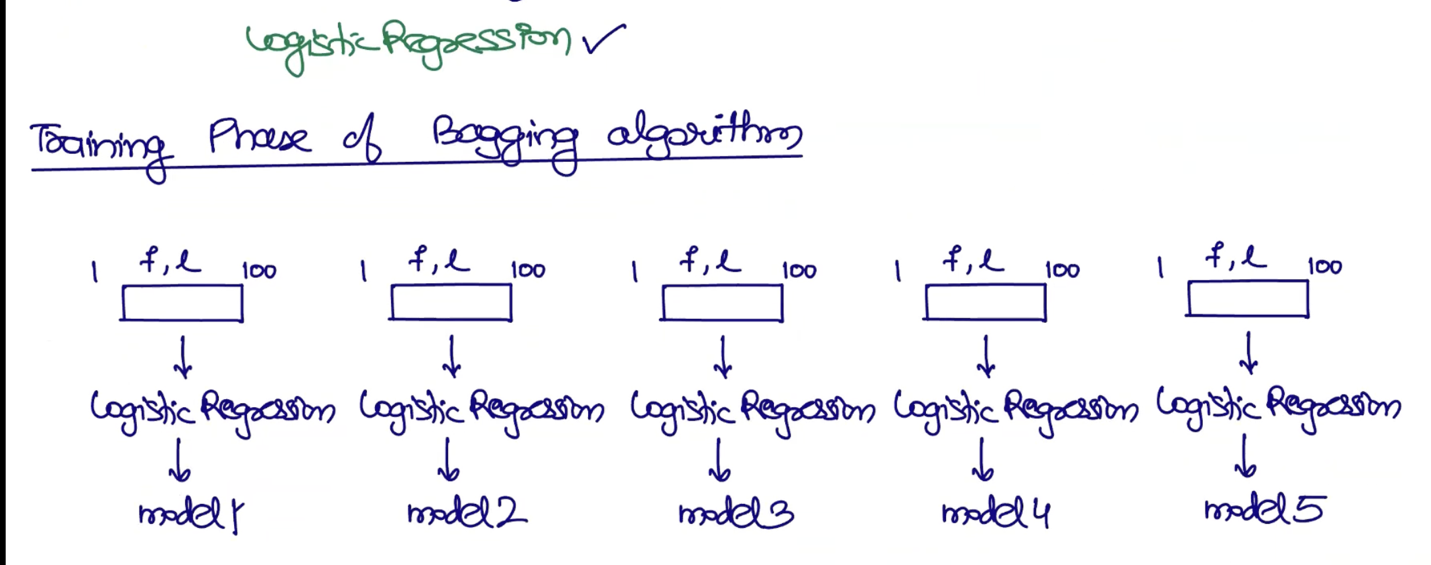
1. How much learners/models you want?

Number\_estimators: 5. [ we are creating 5 weak learners]

1. Which classification algorithm you want to use to create weak learners?

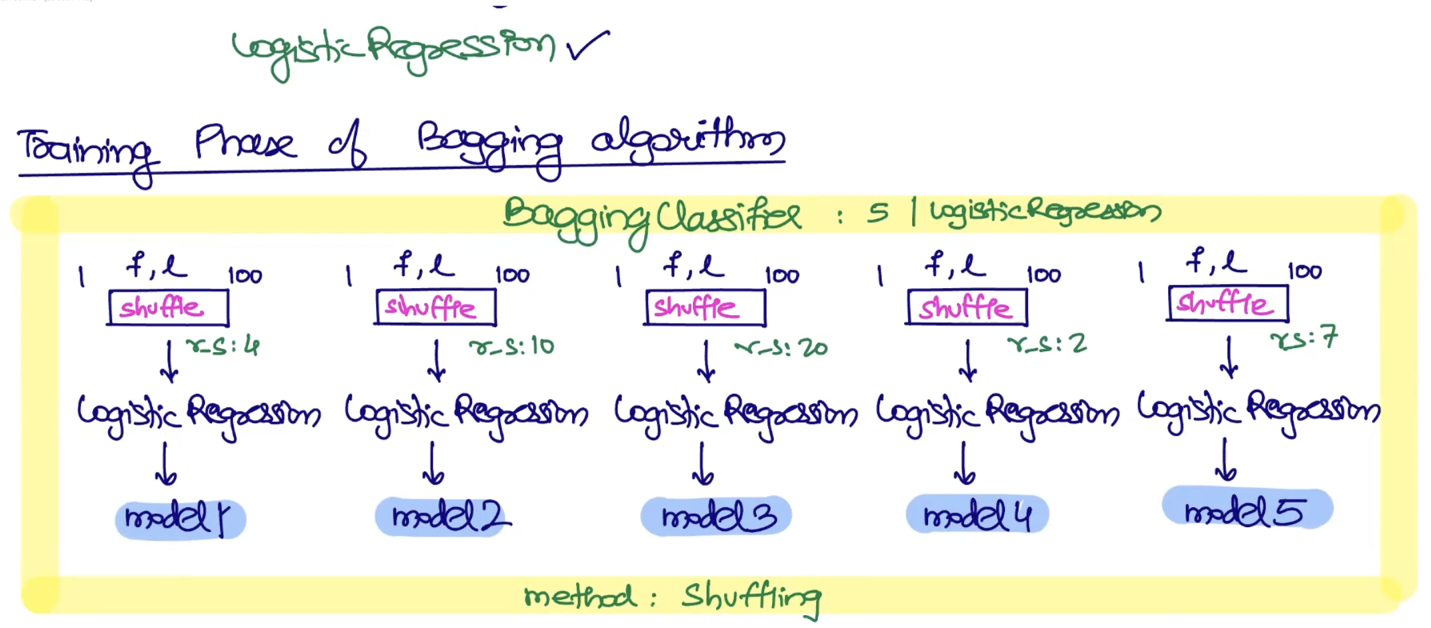
Assume: Logistic Regression

Training Phase of Bagging algorithm:



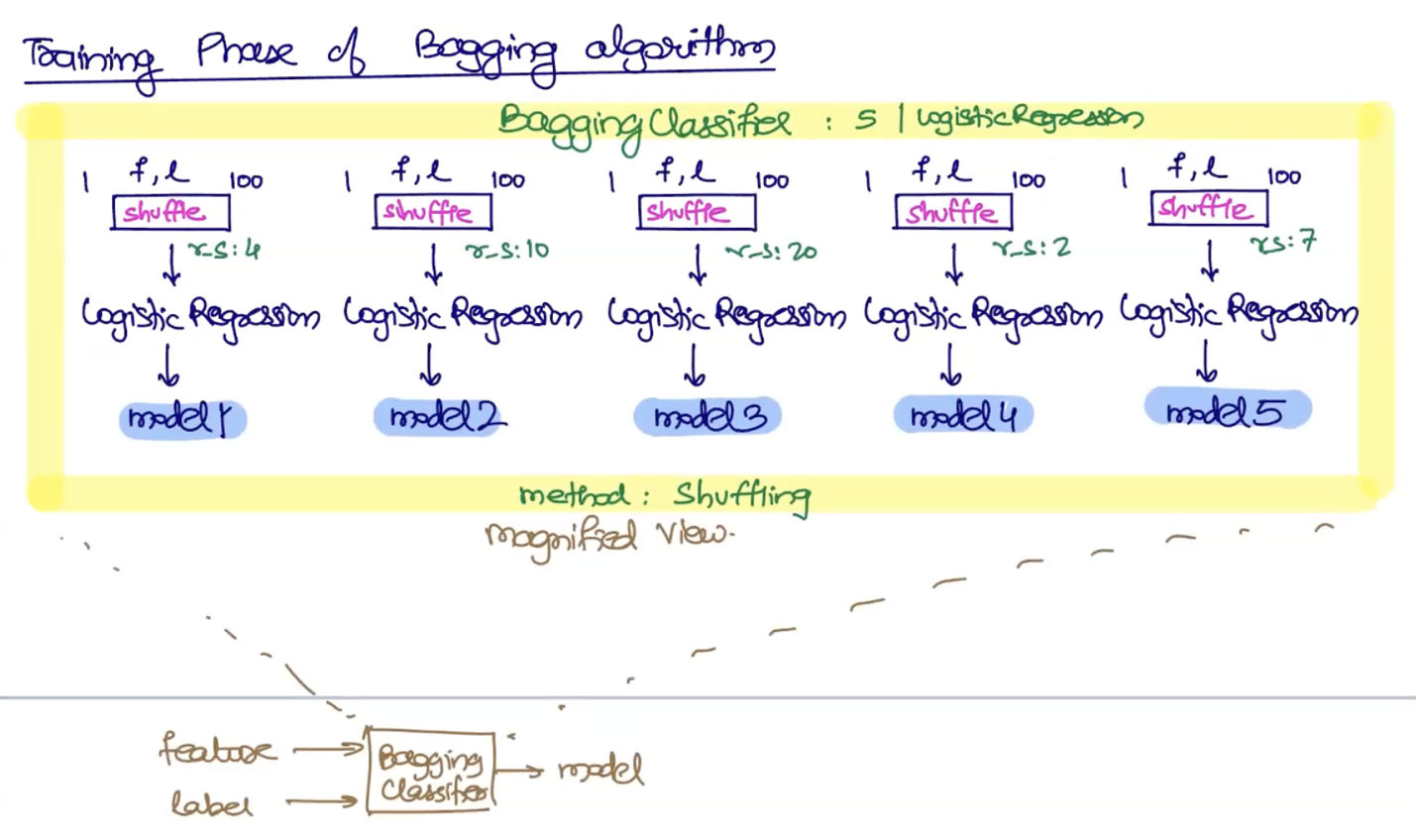
When the dataset is provided to Logistic Regression, bagging algorithm by default will shuffle.

Shuffling means playing with random\_state.



The yellow box is actual single model. Inside this model, you have mini-models. It is the concept of bagging.

How does it work during prediction phase?



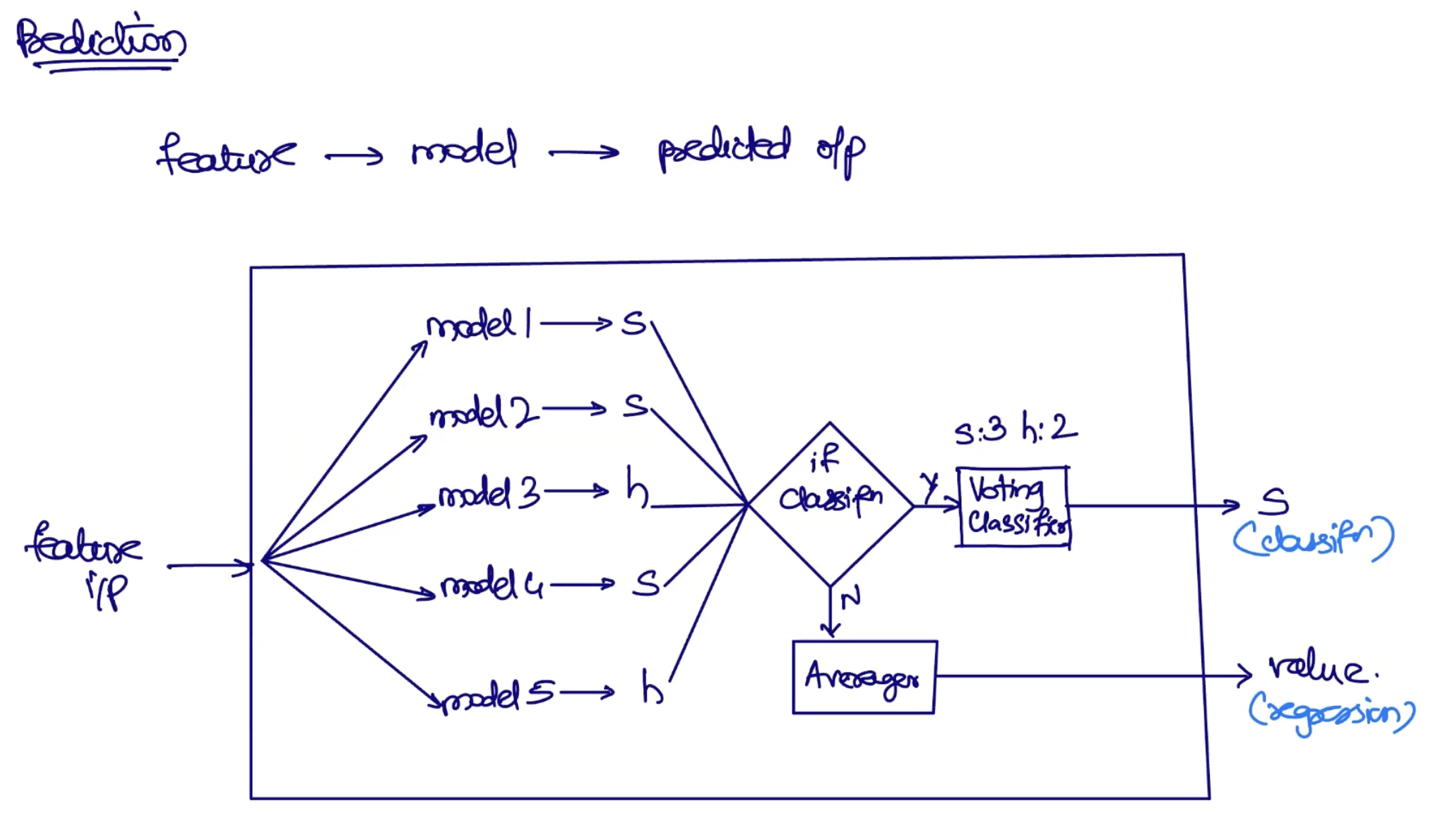
Internally, a set of weak learners are generated. 5 is a hyper parameter.

Prediction:

Feature 🡺 model 🡺 predicted output.

Bagging classifier (bcz it is use case of classification)

Bagging is applicable to supervised-Learning(Regression and classification)



**Training Phase in Bagging algorithms**

Goal is: to train set of weak learners.

There are two method for Training Phase in Bagging algorithms**:**

1. Shuffling method
2. Sampling method

