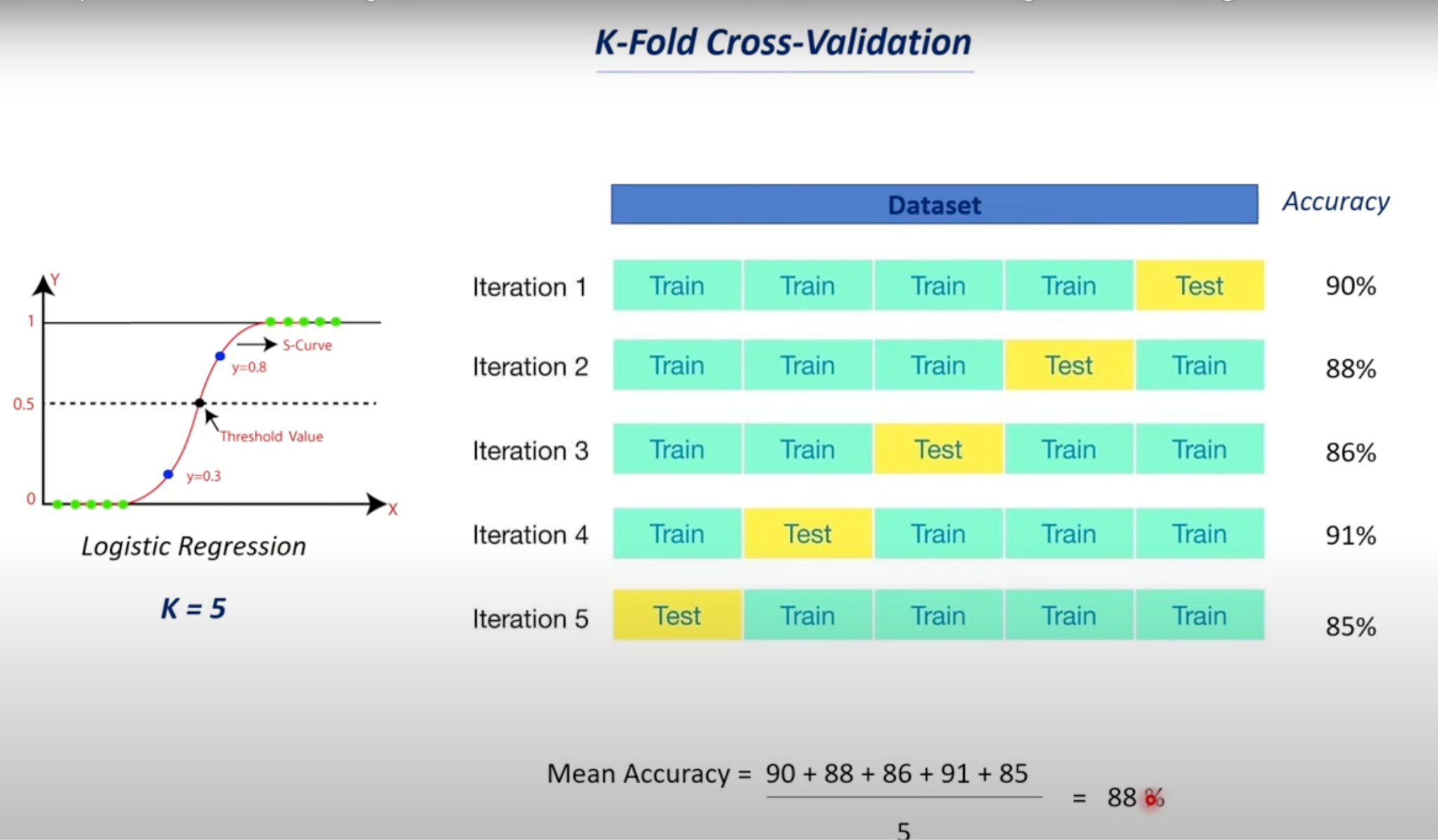
**→ K-Fold Cross-Validation**

1. In this 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.
2. 

→Advantages:

1. Better alternative for train-test-split when the dataset is small
2. Better for multiclass classification problems
3. More reliable
4. Better for model selection.

**→ Hyper Parameter Tuning**

→ parameters - 1) Model Parameters

2) Hyper Parameters

→ Model Parameters:

1. These are the parameters of the model that can be determined by training with training data.
2. There are 2 types - GridsearchCV, RandomizedSearchCV
3. These can be considered as internal parameters.
4. Example - weights and bias

→ Hyper Parameters:

1. These are parameters whose values control the learning process.
2. These are adjustable parameters used to obtain an optimal models.
3. These are considered as external parameters
4. Example - Learning rate, number of iteration

**→ Model Selection**

1. Model selection in machine learning is the process of choosing the best suited model for a particular problem.
2. Selecting a model depends on various factors such aas the dataset, task,nature of the model, etc.

→ Two factors to be considered:

1. Logical reason to select a model
2. Comparing the performance of the model

→ Models can be selected based on:

1. Type of data available:

A- Image & videos - CNN

B- Text data & speech - RNN

C- Numerical data - SVM, Logistic regression, Decision trees, Etc

1. Based on the task we need to carry out:

A- Classification task - SVM, Logistic Regression, Decision Trees, etc

B- Regression task - Linear Regression, Random Forest, Polynomial Regression,etc

C - Clustering Task- K-Means CLustering, Hierarchical clustering, etc

**→ Accuracy Score & Confusion Matrix**

1. Supervised learning - 2 types- classification( prediction a class or discrete value like true or false, 1 or 0) & regression( is about predicting a quantity or continuous value like salary, age, price)
2. Evaluation metric for classification — Accuracy score
3. Evaluation metric for regression — Mean absolute error

→ Accuracy score

1. It is the ratio of the number of correct predictions by total number of prediction times 100.
2. It is not reliable when the dataset has an uneven distribution of classes.

→ COnfusion Matrix

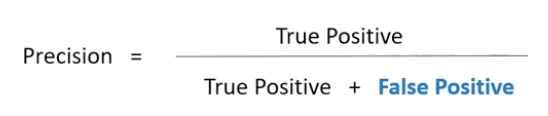
1. It is a matrix used for evaluating the performance of a classification model.
2. It gives us more information than the accuracy score.
3. 

TP + TN = correct prediction

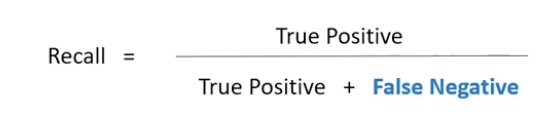
FP + FN = Wrong prediction

Confusion matrix can be used by calling sklearn.matrix.confusion\_matrix

**→ Precision**

1. Precision is the ratio of number of true positive to the total number of predicted potives.
2. It measures, out of the total predicted positive, how many are actually positive.
3. It measure the error caused by false positives, hence it is a good evaluation metric when false positive prediction are critical 

→ **Recall**

1. Recall is the ratio of number of true positives to the total number of actual positive.
2. It measures, out of the total actual positive, how many are predicted as true positive.
3. Recall measures the error caused by the false negatives, hence it is a good evaluation metrc when false negative prediction are crucial

**→ F1 Score**

1. F1 score is an important evaluation metric for binary classification that combines precision & recall.
2. F1 score is the harmonic mean of both precision and recall.
3. This is very useful when the dataset is unbalanced